The University Of Sheffield.

Access to Electronic Thesis

Author:	Yang Li
Thesis title:	Modelling and Adaptive Tracking of Nonstationarities for Both Linear and Nonlinear Systems
Qualification:	PhD

Modelling and Adaptive Tracking of Nonstationarities for Both Linear and Nonlinear Systems

Yang Li

A thesis submitted to the University of Sheffield for the degree of

Doctor of Philosophy

Department of Automatic Control and Systems Engineering

University of Sheffield

Sheffield, UK

Sep. 2011

Abstract

This thesis focuses on the modelling and adaptive tracking problem of both linear and nonlinear time-varying processes. Approaches for the estimation of time-varying parameters can broadly be classified into two categories: the adaptive recursive algorithm methods and the basis function approximation methods. Adaptive algorithms such as block least mean squares (LMS), recursive least squares (RLS) and Kalman filtering, are applied to estimate the time-varying parameters and are capable of tracking the transient variation providing that the variation is slow and smooth. For the basis function method, time-varying parameters are expanded as a finite sequence of predetermined basis functions; the problem of time-varying estimation can then be reduced to a time invariant parameter estimation problem. The basis function expansion approaches are able to track process parameter changes even those with jumps, provided that appropriate basis functions are used. In this thesis, an attractive approach is to expand the time-varying parameters using wavelets as basis functions. Wavelets provide powerful tools for signal processing, with excellent approximation properties and are well suited for approximating general nonstationary signals.

In this work, the application of data-based modelling techniques provides a powerful tool for electrophysiological data modelling and analysis, where a wavelet based modelling approach was applied to model the dynamics of nonstationary signals and capture its transient variations. The work in this thesis contains two parts. The first part deals with the estimation of time-varying linear models in both the time and frequency domains. The performance of tracking and capturing the transient changes of nonstationary systems by using time-varying system identification and modelling in the time-frequency domain has been verified to be an effective approach which outperforms other existing traditional methods such as sliding-window recursive least squares and Kalman filter algorithms. This

technique has been used to investigate and interpret the properties of EEG oscillations of epileptic patients.

The second part deals with the estimation of nonlinear time-varying models, where a novel common model structure selection (CMSS) algorithm has been adapted and extended to identify a robust time-varying common-structured (TVCS) model as a solution to time-varying nonlinear systems identification problems using an online sliding-window approach. The main advantages of the proposed TVCS method are: 1) it produces a less biased or preferably unbiased robust model with better generalisation properties; 2) it enables rapid tracking of transient variations of varying parameters and is more suitable for the estimation of parameters of inherently nonstationary processes. Results on time-varying nonlinear Granger causality analysis have also been investigated to detect and track nonlinear dynamical Granger causalities features.

A main contribution of this thesis is that linear and nonlinear models time-varying modelling techniques have been developed and applied to analyse and interpret multi-frequency signals in both the time and frequency domains. Novel wavelet adaptive tracking algorithms were developed to track both linear and nonlinear system behaviours and the algorithms provide a new tool that can help clinicians interpret EEG signals. In addition, the newly developed methods are generally also applicable to other neuroscience signals. For example, possible applications of our proposed technique could be applied to describe and analyse the coding of speech signals into nerve-action potentials by the inner ear.

Acknowledgements

It is my pleasure to thank so many people who made my thesis possible.

I would like to extend my thanks and sincere gratitude to the Department of Automatic Control and Systems Engineering for giving me permission to commence my thesis in the first instance and to do the necessary research work.

It is very difficult to overstate my deepest respect and gratitude to my supervisor, Prof. Billings S. A. for encouraging me to embark on this research project and for supporting me throughout, and with his enthusiasm, inspiration, knowledge and his great efforts to explain things clearly, simply and patiently. Without his constant encouragement, comments and criticism，my work and knowledge would not be complete. I have learned a lot from Prof. Billings, not just about the research, but about everything else.

It is crucial to thank Dr. Hua-Liang Wei, my second supervisor, patiently and gradually helped me and taught me how to do research, for the many helpful discussions and valuable comments regarding all this work. Also I wish to thank him for his encouragement, support, patience and most importantly fruitful and stimulating discussions and debates about our common interests.

Particularly, this thesis would not have been possible without the financial support from the University of Sheffield and UK Engineering and Physical Sciences Research Council (EPSRC) throughout my graduate program, which is gratefully acknowledged. Without the EPSRC support, it would have not been possible to pursue my dream. It is a pleasure to thank the staff from Department of Automatic Control and Systems Engineering for their enthusiastic support and help.

Special thanks are extended to Dr P. G. Sarrigiannis from Department of Clinical Neurophysiology, Sheffield Teaching Hospitals NHS Foundation, for his stimulating discussions and invaluable advice. The results of this thesis rely on data collected from Sheffield Teaching Hospitals NHS Foundation. Thanks to Dr. Ptolemy for teaching me much of what I know about EEG.

I would also like to thank Prof. Xiao-Feng Liao and Prof. Chuan-Dong Li from Department of Computer Science and Engineering, Chongqing University, Prof. Kwok-Wo Wong from the Department of Electronic Engineering, City University of Hong Kong, and Prof. Heng-Gui Zhang from School of Physics and Astronomy, University of Manchester, respectively, for their constant encouragement and valuable advices throughout my graduate studies. I have learnt a great deal from them. Their insightful suggestion made my work more thorough.

This dissertation would not be possible without help from members of the Department of Automatic Control and Systems Engineering, for example, Dr. L. Z. Guo, Dr. Y. Z. Guo, Dr. Z. Y. Song, Dr. Y. Zhou, Dr. L. Zhao, Dr. W. Yin, Dr. Y. F. Zhao, Bo Pang, Richard Boynton, Peng-Fei Guo, Shu Wang, Yu Liu, Da-Zhi Jiang, Jing-Jing Luo, Xi-Liang Zhang, Xue-Yan Zhao, Xiao-Kai Nie and other research institution staff such as Prof. Colin McCulloch, Dr. Jiang-Hua Liu, Dr. Tao Jiang, Dr. Sheng Xu, Cheng Zhou, Hui Xie, Jie He, and Ben Anderson. Their help with my project was essential to its completion, and they were also always there for me whenever I needed a distraction to keep the endeavour fun.

Last but not least, I am very grateful to my parents, my brother, other family members and in particular my lovely Miss Ni Yang, whose love, constant encouragement and support are backbones of my life. This dissertation is dedicated to them as a gift for the struggles we have gone through over the years. Even though in some moments it was a matter of a bare survival, we have made it. I personally have fulfilled my dream and am grateful to my family for their support.

List of Abbreviations

AIC	Akaike information criterion
ANN	Artificial neural networks
AR	Autoregressive
ARX	Autoregressive with an exogenous
BIC	Bayesian information criterion
CMSS	Common model structure selection
CWT	Continuous wavelet transform
DVV	Delay vector variance
ECG	Electrocardiography
EEG	Electroencephalography
EMD	Empirical mode decomposition
ERR	Error reduction ratio
FB	Fourier–Bessel
FFT	Fast Fourier transform
*f*MRI	Functional magnetic resonance imaging
GCV	Generalized cross-validation
GFRFs	Generalised frequency response functions
HHT	Hilbert Huang transform
HOS	Higher order statistics
HAS	Hilbert spectral analysis
HUP	Heisenberg Uncertainty Principle
*i*AAFT	Iterative amplitude adjusted Fourier transform
IMF	Intrinsic model function
KF	Kalman filtering
LMS	Least mean squares
MAE	Mean absolute error
MDL	Minimum description length

MEG	Magnetoencephalography
MPO	Model predicted output
NARMAX	Nonlinear autoregressive average moving with exogenous input
NPE	Nonlinear prediction error
OLS	Orthogonal least square
OSA	One step ahead prediction
PCA	Principal-component-analysis
PRBS	Pesudo-Random Binary Sequence
PSD	Power spectral density
RBF	Radial basis functions
RLS	Recursive least squares
RMSE	Root mean squared error
RWKF	Random Walk Kalman Filter
SNR	Signal-to-Noise Ratio
S*td*	Standard deviations
STFT	Short-time Fourier transforms
SWRLS	Sliding-window recursive least squares
TIV	Time invariant
TFA	Time frequency analysis
TVARX	Time-varying autoregressive with exogenous
TVCS	Time-varying common structured model
VAR	Vector autoregressive

List of Important Symbols

$\&$	Logic and
$\lceil a \rceil$	Greatest integer greater than or equal to a
$(\cdot)^T$	Transpose
$(\cdot)^*$	Complex conjugate
σ^2, $\mathrm{Var}\{\ \}$	Variance
μ	Forgetting factor
λ	The degree of the polynomial
$\phi_{j,k}(t)$	Wavelet basis function for scale j and shift k
$\{a_i\}_{i=1}^{p}$	$\{a_1, a_2, \cdots, a_p\}$
A	Amplitude of a sinusoidal signal
$\cos(x)$	Cosine function of argument x
$E\{\cdot\}$	Expected value or statistical average of the ensemble
f	Frequency in Hz
f_s	Sample frequency
$h(\tau)$	System unit impulse response
$h_l(\tau_1, \tau_{2,} \cdots, \tau_l)$	l th order impulsive response function
$H_l(f_1, f_2, \cdots, f_l)$	l th order frequency response function
$H_x(t, f)$	Time-frequency representation of $x(t)$ with the
	Frequency expressed in Hz
$\mathrm{Im}\{\cdot\}$	Imaginary part of a complex-valued scalar, vector
$\log(x)$	Natural logarithm of argument x
K	The number of data sets

L	Number of segments
M	The total number of candidate regressors
$\max\{\cdot\}$	Maximum value
$\min\{\cdot\}$	Minimum value
N	Number of time-domain samples
n_θ	Total number of model parameters
n_e	Maximum noise lag
n_u	Maximum input lag
n_y	Maximum output lag
$\mathrm{Re}\{\cdot\}$	Real part of a complex-valued scalar, vector
$\sin(x)$	Sine function of argument x
T_s	Sample interval
$u(t)$	Time-domain input signal
W	Length of sliding window
$x(t)$	Time-domain signal as a function of x
$y(t)$	Time-domain output signal

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Research Background

Parametric model identification of time-varying systems has many applications in diverse engineering fields such as seismic analysis, speech processing and biomedical systems. System identification is the process of detecting model structure and estimating model parameters based on available observation or measured input and output data. Most system identification schemes were developed under the assumption of linear time invariance or stationarity (Johansson 1993; Ljung 1999). Most of the signals encountered e.g. biomedical signals, however, do not meet the stationarity assumptions. Thus, there is a growing interest for dealing with nonstationary signals that arise naturally in these application areas.

There are mainly two classes of approaches to identify and process a time varying (TV) system. Conventionally, the most popular method for dealing with a TV system is to employ an adaptive algorithm such as the recursive least squares (RLS) or Kalman filtering algorithm (Ljung 1999), provided that the system is slowly varying compared to the algorithm' convergence time. For parameters that change fast enough, the adaptive algorithm cannot handle and track systems which vary rapidly. It should be noted that there are three significant deficiencies when the RLS procedure is used to identify TV systems: parameter drift in the case of non-persistent excitation (Sripada & Fisher 1987), varying alertness to system parameter variations, and problems with parameter estimation when using forgetting factors less than 1.

To overcome these aforementioned problems, another common approach for the identification of time-varying systems with quick changing coefficients is to use a basis function expansion that has excellent capability on tracking system variations with time. Many types of basis functions, such as the Legendre polynomial (Niedzwiecki 1988), Fourier series (Pachori & Sircar 2008), Walsh and Haar functions and wavelets (Billings & Wei 2005b; Chon et al. 2005; Tsatsanis & Giannakis 1993; Wei & Billings 2002) are available and capable of representing TV model coefficients. The choices of basis functions have significant effects on the change speeds and smoothness of the estimated parameters. But there is no uniform selection guideline on how to select the appropriate basis functions from the large-family of available basis functions. Wavelet analysis, which has a distinctive property of multiresolution and enables capturing the global as well as local characteristics of TV systems, has been proven a valuable tool for signal processing and successfully applied in various applications including nonlinear signal processing, parametric identification and nonlinear approximation (Billings & Coca 1999; Billings & Wei 2005a; Chon et al. 2005; Chui 1992; Li et al. 2011a; Mallat 1989; Tsatsanis & Giannakis 1993; Wei et al. 2006; Wei et al. 2010; Wei et al. 2008). It is worth stressing that the major question of basis function expansion method is how to select and use only a number of significant basis functions to obtain a parsimonious model of time varying systems and an effective method is always needed to determine how many basis functions should be applied.

Tsatsanis and Giannakis (1993) employed a multiresolution idea for TV system identification by expanding the TV coefficients onto a set of pre-defined linear and nonlinear combinations of wavelet perfect reconstruction filter banks (PRFBs). The F-test and Akaine's information criterion (AIC) approach based on multiresolution procedures were then used to determine the significant terms. The proposed approach was limited to the use of the F-statistics that is dependent on the 'subjectivity' of choosing a threshold. Niedzwiecki (2000) introduced a weighted basis function method to avoid these problems, but the proposed scheme has an

inherent severe deficiency of computational inefficiency in estimating parameters.

Numerous numerical modelling experiences indicated that the first and second order B-splines are non-smooth piecewise functions, which would perform well for the coefficients that change with sharp transients and burst-like spikes over time, whereas the B-splines of higher order would work well on smoothly changing signals. Motivated by this consideration, the B-splines multi-wavelet basis functions as multiresolution wavelet series expansion with different orders for TV model identification are used to capture nonstationary behaviour of biomedical signals such as EEG signals in this thesis. The orthogonal least squares (OLS) procedure and the error reduction ratio (ERR) criteria were then used to select the significant model terms and perform the parametric estimation in the TV expansion (Billings et al. 1989; Chen et al. 1989; Korenberg et al. 1988; Wei & Billings 2002; Wei et al. 2010). Simulation examples and practical applications are given to illustrate the capability and effectiveness of the proposed approach in tracking and capturing various types of nonstationary behaviour of TV systems. Finally, it should be worth stressing that one inherent disadvantage of TV system identification procedure including the adaptive algorithms as well as the basis function approach is that the beginning and endpoints of the data are commonly not accurate. This is a well-documented phenomenon and there are not enough samples in both endpoints of the data.

Biomedical signals such as EEG recordings commonly involve both rapidly and slowly changing behaviours. A novel TV modelling method using a multiwavelet basis function expansion framework is proposed to track and capture the properties of nonstationary EEG signals. Some characteristics of EEG signals and data acquisition procedure are introduced as follows.

Electroencephalography (EEG) recording is an essential clinical tool for the evaluation and treatment of neurophysiologic disorders related to epileptic seizures

that are transient periods of neural activity. It is also widely used as a biomedical signal for the identification and interpretation of different mental states in the human brain. Careful analyses of EEG records can provide valuable insights and improve understanding of the mechanisms causing epileptic disorders. The detection of epileptic discharges in EEG is a significant element in the diagnosis of epilepsy. EEG is the recording of the electrical potentials (activity) produced by the brain, and can be captured by relatively inexpensive equipment. Acquisition procedures are non-invasive and simple. Electrodes are uniformly distributed over the scalp and hooked by wires to a computer; the EEG signals can then be acquired as shown in the flow chart in Figure 1.1. It is worth stressing, in this thesis, that EEG signals or channels shown in Figure 1.1, are generated by taking the difference between measured potentials from two electrodes. For instance, the channel $F_3 - F_7$ is generated by taking the difference between the measured potentials from the electrodes F_3 and F_7. Each EEG channel summarizes localized activity within a region of the brain, for example, the channel $F_3 - F_7$ indicates neural activity originated from the frontal lobe of the left hemisphere.

EEG visual patterns are correlated with functions and dysfunctions of the central neural system, which is regarded as one of the most significant diagnostic tools of neurophysiology. One of most significant applications of EEG technique is to diagnose, classify, synchronize and localize epilepsy, which is a nervous system disorder that produces intense, abnormal electrical activity (spikes) in the brain. Absence seizure is mainly one of the generalized seizures and the underlying pathophysiology is not completely understood. Neurologists make an absence seizure epileptic diagnosis mainly through visual identification of a 3-Hz spike and wave complex (Subasi & Ercelebi 2005). Recently the EEG technique has been the most useful tool for its evaluation. An EEG medical diagnostic framework diagram is briefly described in Figure 1.2. The basic procedure is described as follows: The subjects or patients will be connected to a Jackbox, Electrode montage selector, and

Amplifier. From the amplifier, EEG signal will be acquired, then, some signal processing techniques, for example, notch filtering, will be used to remove some noise and artifacts. Some time-frequency analysis approaches can then be used to analyze and extract the signal features to understand the brain neural activities. The most popular method is typically short time Fourier transform (STFT). In this time-frequency analysis method, there is an issue of time-frequency resolution according to the Heisenberg Uncertainty Principle (HUP). That is, the time-frequency resolution depends on the sliding-window size. If the window size is too narrow, there is a good time resolution, but there is poor frequency resolution. On the contrary, if the window size is too wide, there is a good frequency resolution, but poor time resolution is obtained. Another time-frequency method is the Hilbert Huang transform (HHT) for dealing with the nonstationary and nonlinear signals (Huang et al. 2008). It is also a traditional method to extract the feature of time-frequency resolution from the EEG signal. In this method, there is also a limitation that it only deals with the narrow-band signal. So in this thesis our main goal is to try to find a novel method to extract the signal features with good time-frequency resolution. A novel time-varying parametric modelling method will be investigated to get the good time-frequency resolution from the signal which will be discussed in detail in the following chapter of my thesis. If we can get good time-frequency resolution from the signal, we can then capture and track some more accurate transient information to understand the brain neural activities. Furthermore, these results help the clinician to do some medical diagnosis tasks. Then the analysis diagnosis results will feedback to the subjects or volunteer patients. It should be noted that this diagnostic framework diagram is also suitable for other biomedical signal diagnosis framework.

Electroencephalographers describe the EEG brain activity from the spatial distribution on the scalp including the localized region frontal, posterior, lateral and bilateral as well as the dominant frequency components. The study of different types of rhythmicities of the brain and their relation with different pathologies and

functions has been an important subject of neuroscience, physiology and neurophysiology research (Quiroga 1998). An EEG wave is commonly divided into a wide range of frequency components. However, the range of clinical and physiological interest is between 0.5 and 30 Hz. This range is divided into a number of frequency band components (Cohen 1986; Cohen 2000) shown in Table 1.1. Most of the cerebral oscillation observed in the scalp EEG is distributed in the range of 1–20 Hz. For instance, the predominant physiological *Alpha* frequency component appears most prominently on posterior channels during the relaxed state of healthy subjects with their eyes closed. Frequency activity below or above the range of 1 – 20 Hz probably is contaminated by an artefact from clinical recording techniques. Some EEG artifacts have specific activity and scalp topography which can be more identified in the frequency domain. EEG signals contaminated by artifacts with very low amplitudes normally require more rigorous recording technique to interpret the EEG signals easily. For example, the power line frequency in Europe is 50 Hz, while in North America it is 60 Hz. Some signal processing techniques are applied to remove the artifacts prior to analysis and applications of the EEG signals.

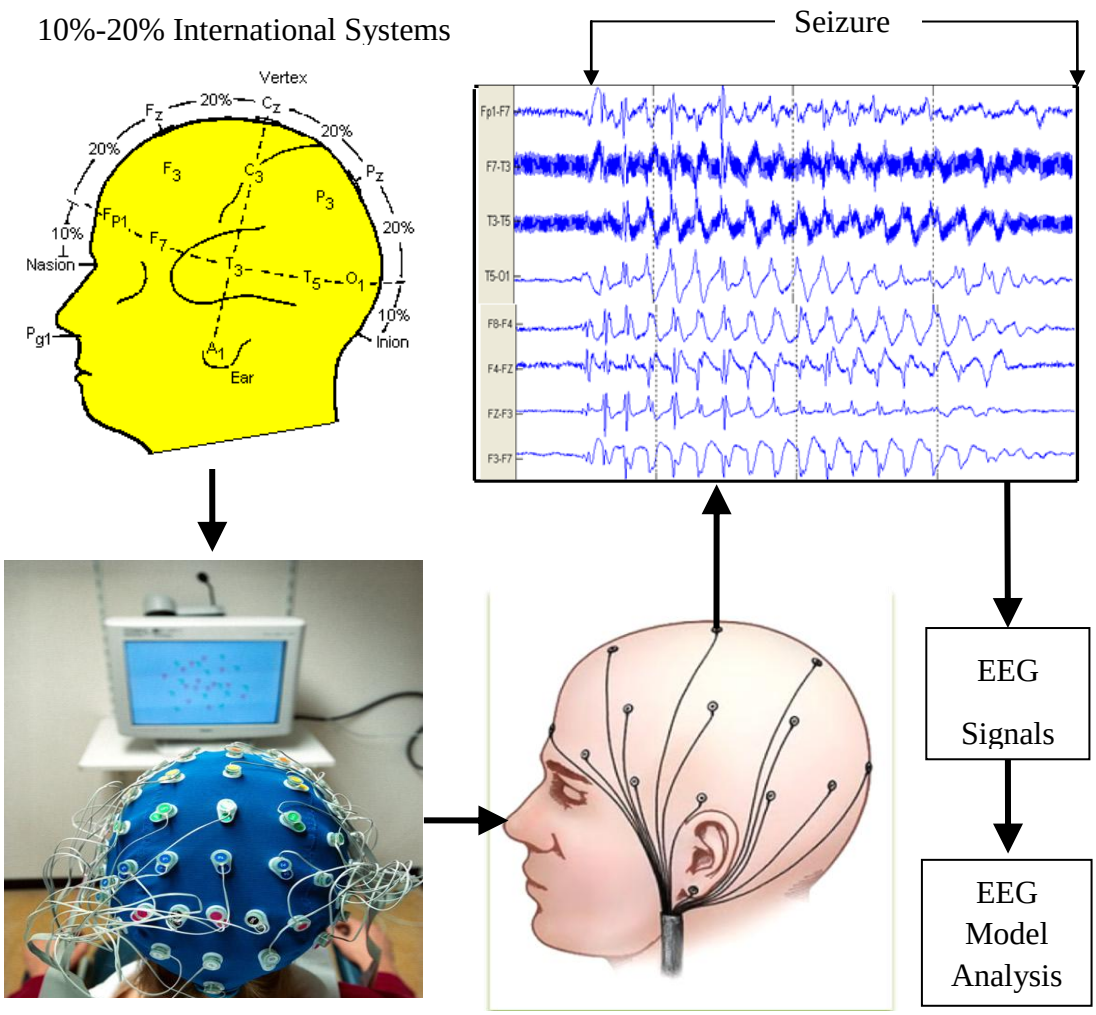


Figure 1.1 The flow chart of EEG acquisition from the locations of 32-channel surface electrodes placed on the brain cortex scalp on the basis of a 10% or 20% of a measured length from a known landmark on the skull. Notation: F represents frontal; T, temporal, and O, occipital.

Photos: http://www.lsa.umich.edu/psych/danielweissmanlab/images/eegimage.jpg

http://hipnozvehipnoterapi.com/images/bn7_eeg.jpg

http://faculty.washington.edu/chudler/1020side.gif

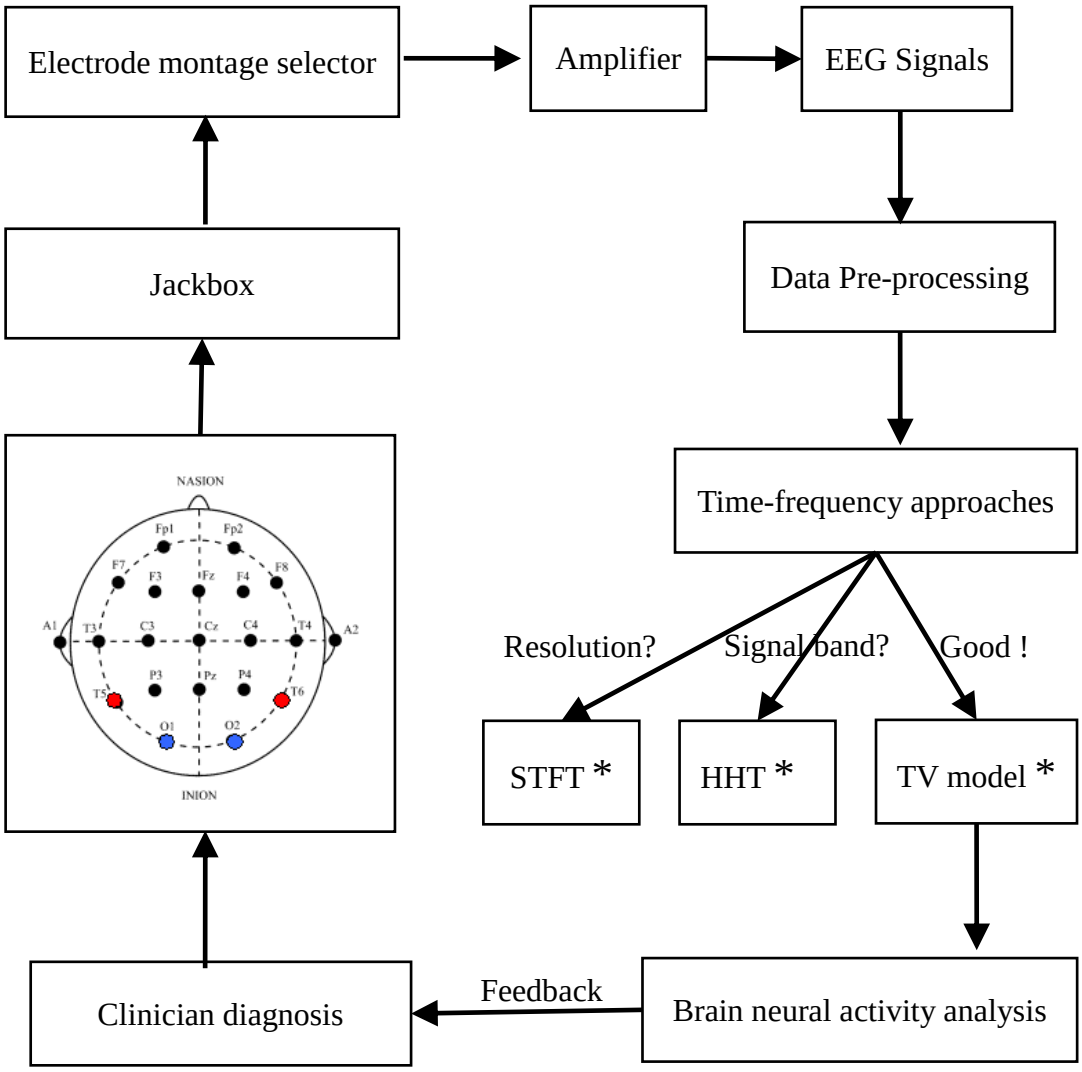


Figure 1.2 The EEG medical diagnostic framework diagram. Notation: * STFT indicates Short Time Fourier transform; HHT: Hilbert Huang transform, TV: time-varying.

Photo: http://filip.fd.cvut.cz/obrazky/112_01_05.gif

Table 1.1 Comparison of EEG frequency components.

Type	Frequency (Hz)	Location and Pathology
Delta	0.5 – 3.5	frontally in adults, posteriorly in children; locations are correlated with different pathologies such as metabolic encephalopathy hydrocephalus.
Theta	3.5 – 7.5	found in locations not related to task at hand, associated with inhibition of elicited response; pathology: metabolic encephalopathy and deep midline disorders.
Alpha	7.5 – 12.5	posterior regions of head, most pronounced in occipital and posterior locations, both sides; pathology: coma.
Beta	12.5 – 30	best defined in central and frontal locations; both sides, symmetrical distribution; pathology: benzodiazepines.
Gamma	30 – 60	somatosensory cortex; pathology: in gamma band activity, a decrease may be associated with cognitive decline.

Cohen (1986). *Biomedical Signal Processing*.

1.2 Motivations and Goals of the Project

Signal processing techniques are commonly used for feature extraction in medical diagnosis, fault detection and many other application fields. The main purpose of signal processing is to reveal underlying information on specific problems in these applications. Signal processing techniques based on modern system identification algorithms can be classified as time, frequency, and time-frequency domain methods. In real-world applications and signal processing, a signal is normally assumed to be stationary. However, in many practical applications such as electrophysiological data this assumption is not always accurate, because relevant statistical characteristics change over the time course, depending on the mental states that are active at any given time instant. Signals with time-varying frequency components are highly nonstationary. Modelling of nonstationary signals is very difficult and reliable parametric models normally do not exist. In practice most of the signals encountered cannot satisfy the stationary assumption conditions, which explains the growing interest in nonstationary signal processing and applications. Time-frequency analysis (TFA) of nonstationary signals is of great interest and is very significant since the time history of frequency is a very powerful method in signal characterization. The TFA approaches are most popularly performed by STFT, wavelet and HHT. The main drawbacks of these approaches have been briefly described in section 1.1.

This thesis was motivated by the desire to take advantage of linear and nonlinear system identification techniques and some estimation algorithms to obtain the time-varying linear and nonlinear models of the biomedical signals that can further capture the dynamics of the signals, and extract the time-frequency feature analysis. Good frequency resolution should be possible by means of autoregressive (AR) or autoregressive moving average (ARMA) parametric models. The parametric method of spectral estimation for stationary signals has been investigated (Kay

1988). The spectral estimation results have shown that, even if the available signal is very short, the parametric method can yield good high frequency resolution. Therefore, this parametric approach also has been adapted so that it can be extended to spectral estimation of nonstationary signals. The time-frequency analysis scheme encompasses signal processing approaches for nonstationary signals. Various methods are proposed to resolve such a task. Considering the adaptive parametric model method as a feature in the time-frequency domain, the approaches for feature extraction are diverse and range from simple, such as the adaptive RLS algorithm, to more complex, such as basis function expansion approach. These methods are applied since they can provide more accurate results than the traditional STFT method.

If the parametric model of nonstationary signals for both linear and nonlinear models in the time-frequency domain is applied as a feature for extraction or recognition, the following questions have to be answered in this thesis. For instance, how can we select the model structure and parameter estimation? More generally, is it possible to improve the accuracy of the time-varying estimation and enhance the adaptive tracking capability of nonstationary signals using linear and nonlinear models? Also, if the rapid tracking of estimated time-varying parameters is used as a feature in a feature extraction process, how can we carry out the time-frequency resolution map in a certain time-frequency band? Should the existing adaptive algorithm techniques be used? Or should new basis function expansion schemes for time-varying parameters estimation be developed which depend on the existing algorithms? How can we improve the robustness of identified models based on the orthogonal least squares algorithm (OLS)? In particular, if the time-varying parametric model identified is used to analyse the EEG signals of causal influence between different channels, what are the advantages for the proposed time-varying model compared to the existing linear and nonlinear parametric model approaches? The goal of this thesis is to provide some answers to these questions.

1.3 Organization of the Thesis

The main purpose of the thesis deals with the application of system identification techniques to model the dynamic relationships and capture the transient variations of nonstationary signals adaptively between EEG signals. The thesis is divided into two parts. The first part deals with the estimation of time-varying linear models in both time and frequency domain, and the second part deals with the estimation of nonlinear EEG signal models that provide a substitute for the family of linear models estimated previously. Nonlinear model estimation approaches and some applications are discussed in the final stage of the thesis.

The thesis is organized as follows.

The main body of the thesis starts in Chapter §2 with a literature review of the parametric modelling and time-frequency domain analysis. Chapter §2 provides an in-depth analysis and comparison of the existing adaptive algorithms for estimating the time-varying parametric model. Some time-frequency analysis techniques are also discussed for extraction of the considered feature. This literature review attempts to consider and review the issue from different angles, and it mainly provides an overview of contributions in the following three subfields:

- Design of new time-frequency analysis of nonstationary signals;
- Robust algorithm of nonlinear system identification;
- Framework of Granger causality for nonlinear models.

As a special analysis case, Chapter §2 comments on the transient frequency estimation on the basis of the time-frequency analysis approach for nonstationary signals. The estimation procedure considered is based on a multi-wavelet basis function expansion scheme. In addition, some of the existing problems or

disadvantages in the parametric estimation field are also discussed. The problems are presented, and their significance is further explained. Time-frequency analysis with high frequency resolution of nonstationary signals is defined by using the parametric method with time-varying coefficients. Information about the nonlinear dynamics is introduced by analysing the estimated models in the frequency domain by calculating their *generalised frequency response functions* (GFRFs). The detection of linear and nonlinear causality between signals is also presented and reviewed briefly.

Chapter §3 introduces the fundamental concepts and model structure relating to the modelling and parametric estimation of linear time varying systems in the time domain. A new time-varying parametric modelling approach has been developed, where the associated time-dependent coefficients are approximated using the multi-wavelet basis functions including the cardinal B-splines basis functions. One advantage of the proposed approach is that it can be used to track rapidly or even sharply varying processes, while traditional adaptive algorithms such as the LMS and RLS approaches cannot track the system's time evolution if the coefficients change fast enough. Another advantage is that the proposed approach can track rapid time variation and is more suitable for the estimation of process parameters of inherently nonstationary processes. Two examples, one for a simulation signal, and another of a mechanical system, are given to show the effectiveness and applicability of the new TVARX modelling method. So we advocate the use of a multi-wavelet basis functions because of its flexibility in capturing the signal's characteristics at different scales.

Chapter §4 introduces a novel time varying model identification scheme based on multiwavelet basis functions. A single input/output TVARX model was used to illustrate the model structure and an orthogonal least squares (OLS) algorithm was applied to estimate multi-wavelet coefficients. A time-dependent spectrum function was defined, which is very useful to accurately characterize the frequency-domain

properties of the EEG signals. The numerical properties of the proposed time-varying modelling method are investigated by using a set of test signals. The simulation results showed the proposed approach outperforms the traditional adaptive algorithms such as the LMS and RLS method. In addition, two illustrative practical examples with application to real EEG data are considered. The TVARX model can capture well the rapid dynamics of EEG data. Notably the transient time-dependent spectrum estimated from the model shows very interesting activities from different frequency bands.

Chapter §5 presents a novel common model structure selection (CMSS) algorithm, which can be used to identify a robust time-varying common- structured (TVCS) model. Once the common-structured model has been determined, relevant time-varying model parameters can then be estimated using a SWRLS algorithm. The proposed approach in this Chapter has mainly two advantages which are described as below.

- First, the TVCS model can produce less biased or preferably an unbiased robust model with better generalisation properties. The ‘hold-out’ or ‘split-sample’ data partitioning methods are commonly used in identification, where the available observed data are conventionally partitioned into two parts: the training data that are used for model identification and the test data that are used for model performance validation. Generally, the hold-out data partitioning approach is very convenient and the associated model identification procedure is easy to implement. However, it should be noted that the resultant model obtained from such a once-partitioned single training dataset may occasionally lack robustness and generalisation to represent future unseen data, because the performance of the identified model may be highly dependent on how the data partition is made. This problem will be even more exaggerated when the system is time varying. To overcome the drawback of the hold-out data partitioning approach, in this Chapter, a new common model

structure selection (CMSS) algorithm is proposed to produce less biased or preferably unbiased robust models. For the case where the sliding-window size is equal to the available observed data length, the new CMSS approach will simplify to the case of the hold-out data partitioning method.

- Another advantage is that the approach can be used to track rapid changes and capture transient variations of varying parameters and is more suitable for the estimation of process parameters of inherently nonstationary processes. Two examples, one based on simulation data and the other using a real EEG signal, are given to show the effectiveness and applicability of the new TVCS modelling method using an online sliding-window approach. We can fit individual models to each window, but there are many applications such as fault detection, biomedical engineering where the underlying system characteristics can be revealed by one common model over a series of windows.

This is a more challenging problem and is addressed in this Chapter to achieve a parsimonious identified TVCS model. The TVCS model is different from the traditional multi-input and multi-output (MIMO) model structure, where each subsystem model may not need to share the same common model structure and which often involves one single data set.

Chapter §6 investigates Granger causality, which is a fundamental tool for the description of the causal interaction of two signals. In this Chapter, a novel linear and nonlinear time-varying parametric modelling and identification approach using data-driven methods is proposed for the adaptive estimation of nonstationary EEG Granger causality processes to detect the transient dynamic causal directional interactions between EEG signals. The time-varying model proposed allows identification of the direction of information flow between brain areas, extending the Granger causality concept to transient time-varying processes. A numerical

example demonstrates a good performance of the time-varying Granger causality of detecting transient dynamical causal relations over the time course. This approach is applied to analyse EEG signals to track and detect the causal influences between EEG signals. One advantage of the proposed model is that our results can be more interpretable and yield new insights into the transient directed dynamical Granger causality interactions.

As a summary, the main contribution of this thesis and the suggestions for future studies are given in Chapter §7.

Chapter 2

Literature Review

2.1 Introduction

Nonstationary signals and time-varying systems are often applied to numerous applications including speech recognition, process control, fault detection, medical diagnosis, biomedical signal processing and many other fields, where the underlying time-varying systems are subject to fast and abrupt changing environments (Wellstead & Zarrop 1991). The primary objective of time-varying signal processing in these applications is to provide underlying information on specific problems. These techniques can be classified either as time, frequency, or time-frequency domain based on different algorithms. In general, adaptive algorithms have been used to track the system's variations provided that the system change is slow in comparison with the algorithm's convergence time. Parametric identification algorithms and signal processing can be used for feature extraction if an accurate model of the signal exists in a selected representation space (Manolakis et al. 2005). The models may be used to analyse the time-varying signals, leading to an estimation of the power spectral density (PSD) of the signal, and subsequently a recovered signal from the identified model having the same spectral characteristics as the original one may be synthesized. This is the reason that the time-varying parametric approach has received much attention for speech processing as well as for various other electrophysiological fields such as ECG, EEG, and seismography. However, such modelling techniques have some limitations as well, for example, a strong limitation of these techniques needs to assume a stationary signal. In practical analysis, one way to relax this limitation is to perform the identification of the model over short segments; however, this

requires a compromise between the accuracy that can be achieved with a short data segment and the faithfulness with which the spectrum must be followed. This is one reason why parametric methods need to be proposed for nonstationary signals. Modelling of nonstationary signals is more difficult and reliable parametric models often do not exist, except in very few special cases, e.g. multi component chirp signals (Mukhopadhyay & Sircar 1997). Most of the signals encountered in practice do not satisfy the stationary conditions (Cohen 1995), which explain the growing interest in nonstationary signal processing.

Several approaches have been proposed to deal with the time-varying modelling problems (Bouzeghoub et al. 2000). One of the most popular approaches for identification of time-varying systems is to adopt an adaptive algorithm provided that the time variations are slow so that the system changes can be tracked. The adaptive recursive estimation methods are a stochastic approach, where the coefficients of the associated models are regarded as random processes with some stochastic model structure, the most popular methods to deal with this class of random models include the LMS, RLS and Kalman filtering (KF) algorithms. It should be noted that a potential problem using the Kalman filtering for parameter estimation assumes an appropriate model for the parameter trajectories (Hayes 1996; Tsatsanis & Giannakis 1993). Opposite to the adaptive algorithms, the basis function expansion and regression approach for parametric identification of linear and nonlinear time-varying systems is a deterministic parametric modelling method, where the associated time-varying coefficients can be expanded using pre-defined finite basis functions. Here these coefficients are expressed using a linear or nonlinear combination of a finite number of basis functions. The time-varying identification problem then becomes time invariant with respect to the parameters in the expansions. Hence, the initial time-varying modelling problem can be simplified to deterministic regression selection and parameter estimation. The two main basis problems encountered when this approach is applied to general time-varying parametric modelling problems include how to choose the basis

functions and how to select the significant ones from the family of the basis functions. If these issues can be solved, the final model can be expressed using these 'significant' basis functions.

Parametric spectral estimation has received much attention over the last several years. The main advantages of parametric spectral estimation are an improved accuracy at high signal-to-noise ratios, especially for short data samples, and the flexibility of analysis. This has led to a growing interest in nonstationary signal processing including time-frequency distribution and time-varying parametric spectrum estimation methods. In contrast with most non-parametric methods of nonstationary spectrum estimation including narrow-band filtering, short-time Fourier transforms (STFT), Wigner distribution (Wigner 1932) and several transformations leading to time-frequency representations which are relatively well developed, alternative parsimonious descriptions can be adopted in cases where the signal can be represented by a time-varying parametric spectral estimator.

2.2 Adaptive algorithms

The time-varying model identification is commonly utilized to represent an input-output relationship of time-varying linear and nonlinear systems from the dynamic responses and input forces. The adaptive recursive algorithms such as LMS and RLS are one of the most popular techniques to estimate time-dependent coefficients of the time-varying linear and nonlinear model.

2.2.1 Least mean squares and Block least mean squares algorithms

The LMS algorithm, which is an online approach, was well described by (Hayes

1996). The LMS algorithm is an important stochastic gradient algorithm. A significant feature of the LMS algorithm is its simplicity, which does not require matrix inversion, nor does it require measurements of the pertinent correlation functions. The LMS algorithm is simple compared to other adaptive recursive algorithms such as RLS algorithms. In spite of the computational efficiency of the LMS algorithm, additional simplifications may be necessary in some applications including high speed digital communication and biomedical applications. However, several modified approaches of LMS algorithms may be applied to reduce the computational requirements. One of the most popular of these is the block LMS algorithm, which is identical to the LMS algorithm except that the TV coefficients in the mathematical model that needs to be identified are updated only once for each block of L samples (Clark et al. 1981). For example, the TV coefficients are held constant over each block of L samples, and the desired output $\hat{d}(n)$ and the estimation error $e(n)$ for each value of sample index n within the block are calculated using the coefficients for each block. At the end of each block, the coefficients are then updated using an average of the L gradient estimates over the block. The procedure of the block LMS algorithm is given in Chapter 3 in detail. It should be noted that some other simplification of the LMS algorithms, such as the sign algorithm where the LMS coefficient update equation is modified by applying the sign operator to both of the error or the data, variable step-size algorithm, gradient adaptive lattice filter that the gradient adaptive lattice has some important advantages over the LMS algorithm structure, have discussed in (Hayes 1996).

2.2.2 Recursive least squares and sliding-window recursive least squares

The RLS algorithm, which is also a traditional online approach, was well discussed

by (Ljung 1999). Although the RLS method has high computational efficiency when estimating TV parameters, the traditional RLS algorithm can show slow tracking capability for time-varying coefficients and is highly sensitive to the additive noise and initial conditions. To improve these shortcomings, variable forgetting factors (Fortescue et al. 1981; Leung & So 2005; Toplis & Pasupathy 1988), covariance matrix resetting (Jiang & Cook 1992; Park & Jun 1992), the sliding window techniques (Belge & Miller 2000; Choi & Bien 1989) have been incorporated into the RLS algorithm to improve the tracking ability for TV systems. Compared to exponentially weighted RLS, the sliding window RLS requires about twice the number of multiplications and additions and additionally requires the initial values to be stored, which storage requirement may potentially be a problem for long windows.

2.2.3 Comparison with least means squares and recursive least squares

In this subsection, the technique performances of LMS algorithm and RLS approach for processing nonstationary signals will be discussed. These techniques have been extensively applied to a variety of applications such as signal modelling, nonlinear approximation, spectrum estimation, system identification, adaptive equalization and noise cancellation. The LMS algorithm, which is simple and often effective, does not require any ensemble averages to be known. Several modifications of the LMS algorithm then have been discussed. For example, the normalized LMS algorithm can simplify the selection of the step size to ensure that the coefficients converge. The leaky LMS algorithm is useful in overcoming the problems that occur when the autocorrelation matrix of the input process is singular. The block LMS algorithm, which is designed to increase the efficiency of the LMS algorithm, has also been discussed in the literature. In the block LMS algorithm, the filter coefficients are held constant over blocks of length L, which allows for

the use of fast convolution algorithms to calculate the filter output. The lattice adaptive algorithm converges more rapidly than the LMS adaptive approach, and tends to be less sensitive to the eigenvalue spread in the autocorrelation matrix of input signals. The RLS algorithm minimizes a deterministic least squares error. It could be that, since the RLS approach is recursive, it is possible for the TV coefficients estimations to become unstable. As a result, adaptive recursive approaches are not as widely used in applications as LMS approaches. The exponentially weighted RLS algorithm and the sliding window RLS algorithm have been given, although computationally more complex than the LMS adaptive algorithm, for wide-sense stationary processes the exponentially weighted RLS algorithm convergences much more rapidly. The adaptive lattice algorithm for recursive filters presented in (Fallah & Ramachandran 1992; Parikh et al. 1980) indicated that it could be made to remain stable and convergence fast during the adaptation process while the computational complexity is encountered. However, in order to track a nonstationary process effectively, it is necessary to use either the exponentially weighted RLS algorithm or the sliding window RLS approach.

The treatment of adaptive algorithm in this chapter is, by no means, complete. Many other adaptive approaches have been developed and many papers have been published that analyze the performance of adaptive algorithms and evaluate their effectiveness in different applications. Some notable omissions include the fast RLS recursive algorithm, the RLS lattice algorithm, and the nonlinear adaptive algorithm (Zheng & Lin 2003).

Generally speaking, the RLS algorithm approach is less sensitive to eigenvalue disparities in the autocorrelation matrix of the input signal for stationary processes than LMS. On the other hand, the RLS algorithm does not perform very well in tracking nonstationary processes without exponential weighting. This is the reason that, with forgetting factor to be equal 1, all of the data is equally weighted in estimating the correlations. Note that, although exponential weighting improves the

tracking characteristics of RLS, there is no a selection guideline on how to choose the suitable forgetting factor, hence in some cases, the LMS algorithm may have better tracking properties. For a more detail discussion of the RLS algorithm and the LMS algorithm, the readers can refer to the literatures (Haykin 2003; Ljung 1999).

2.3 Time-frequency analysis approaches

In this section, there are two different approaches, Fast Fourier Transform (FFT)-based methods (called non-parametric methods), and model-based methods (also called parametric methods), which are discussed.

2.3.1 Fast Fourier transform (FFT)-based methods

The classical methods make no assumption about how the data were generated and hence are called nonparametric. Although the spectral estimates are expressed as a function of the continuous frequency variable f, in practice, the estimates are computed at discrete frequencies via the FFT algorithm. The periodogram and Welch method are the FFT-based methods. The Welch spectral estimator can be efficiently computed via FFT and is one of the most frequently used PSD estimation methods. In the Welch method, signals are divided into overlapping segments, each data segment is windowed, periodograms are calculated and then the average of the periodograms is found. The Welch spectral estimator is defined as

$$\hat{P}_i(f) = \frac{1}{MU}\left|\sum_{n=0}^{M-1} x_i(n)w(n)\exp(-j2\pi fn/f_s)\right|^2,\ i = 0,1,\cdots,K-1,$$

and

$$\hat{P}_W(f) = \frac{1}{K}\sum_{i=0}^{K-1}\hat{P}_i(f), \tag{2.1}$$

where $\hat{P}_i(f)$ is the periodogram estimate of each signal interval, $w(n)$ is the data window, U is the normalization factor for the power in the window function $w(n)$ given as $U = \frac{1}{M}\sum_{n=0}^{M-1} w^2(n)$, and f_s is the sampling frequency. $\hat{P}_W(f)$ is the Welch PSD estimate, M is the length of each signal interval and K is the number of signal interval (Proakis & Manolakis 1996).

2.3.2 Model-based methods

Model-based methods for spectral estimation consist of choosing an appropriate model, estimating the parameters of the model, and then substituting these estimated values into the theoretical PSD expressions. The models discussed are rational transfer function models. They include the autoregressive (AR) model, and the autoregressive with an exogenous (ARX) model. In the Yule-walker method, PSD estimation is formed as (Proakis & Manolakis 1996; Subasi 2007):

$$\hat{\mathrm{P}}_{\mathrm{YW}}(f) = \frac{\hat{\sigma}_{wp}^2}{\left|1+\sum_{k=1}^{P}\hat{a}(k)e^{-j2\pi fk/f_s}\right|^2}, \tag{2.2}$$

where $\hat{\sigma}_{wp}^2$ is the estimated minimum mean-square value for the p th order predictor, $\hat{a}(k)$ are the AR model parameters estimation, and f_s is the sampling frequency, respectively. The Burg method is based on minimization of the forward and backward predicted errors and estimation of the reflection coefficient. From the estimates of the AR parameters, PSD estimation is formed as (Proakis & Manolakis 1996; Subasi 2007)

$$\hat{\mathrm{P}}_{\mathrm{BURG}}\left(f\right)=\frac{\hat{E}_P}{\left|1+\sum_{k=1}^{P}\hat{a}_P\left(k\right)e^{-j2\pi fk/f_s}\right|^2}, \tag{2.3}$$

where $\hat{E}_p^f$ and $\hat{E}_p^b$ indicate the least-squares estimate of the forward and backward errors, $\hat{E}_P=\hat{E}_p^f+\hat{E}_p^b$ is the total least squares error, respectively.

2.3.3 Time-frequency methods based on wavelets

The electrophysiological signals including EEG signal are TV and have highly complex time-frequency characteristics. In practice, the stationary condition for the TV signals can be satisfied by dividing the signal into blocks of short segments in which the signal segment can be assumed to be stationary. This method is referred to as the STFT that is very typically time-frequency analysis approach. However, there exists an issue of time-frequency resolution according to the Heisenberg Uncertainty Principle. Namely, the problem with the STFT is the length of desired segment and the time-frequency resolution depends on the sliding-window size. Choosing a short analysis window may lead to poor frequency resolution. On the other hand, a long analysis window may improve the frequency resolution but compromises the assumption of stationarity within the window. The most popular time-frequency analysis method for biomedical signal study may be the continuous wavelet transform (CWT). The advantage of the CWT is that it compromises the time-frequency resolution tradeoff problem using a short window at high frequencies and a wide window at low frequencies. However, the CWT has a degraded frequency resolution for high-frequency components and a degraded time resolution for low-frequency components. Herein, the CWT approach, which is a nonparametric time-frequency analysis approach, maybe cannot precisely identify and acquire temporal and spectral information for fast variations of electrophysiological signals. Another time-frequency analysis method is the Hilbert Huang transform (HHT) for dealing with the nonstationary and nonlinear signals. It

should be noted that the HHT approach is an empirical approach that use empirical mode decomposition (EMD) method to decompose a signal into intrinsic model function (IMF) components (Huang et al. 1998; Yang et al. 2003), where the instantaneous frequency can be calculated using the Hilbert Transform, and then a Hilbert spectral analysis (HSA) approach is used to obtain instantaneous frequency data from each IMF component. However, it should been noted that the EMD is only limited to narrow-band signals for distinguishing different components. The narrow band discussed here may include components either having adjacent frequencies or not being adjacent in frequency but where one of the components has a much higher energy intensity than the other components.

An alternative way to analyze the time-varying biomedical signals to obtain the good time-frequency resolution is the time-dependent spectral estimation method: The p th order time-varying AR model, TVAR(p), is formulated as below

$$y(t)=\sum_{i=1}^{p} a_i(t)y(t-i)+e(t), \tag{2.4}$$

and the time-varying ARX (TVARX(p,q)) is given by

$$y(t)=\sum_{i=1}^{p} a_i(t)y(t-i)+\sum_{l=0}^{q} b_{\,l}(t)u(t-l)+e(t), \tag{2.5}$$

where t is the time instant or sampling index of the signal $y(t)$, and the input signal $u(t)$ and the output signal $y(t)$ are measurable; $a_i(t)$ and $b_l(t)$ are the time-varying ARX(p,q) model parameters to be determined; p and q are the model orders for the output and input, respectively. It is assumed that the model orders (maximum lags) are time invariant. The term $e(t)$ represents the model residual or modelling error that can often be treated as a stationary white noise sequence with zero mean and variance σ_e^2, $a_i(t)$ and $b_l(t)$ are the TV coefficients.

The proposed method is to expand the TV parameters $a_i(t)$ and $b_l(t)$ onto multi-wavelet basis function $\pi_m(t)$ for $m=1,2,\cdots,R.$ such that the following expressions hold:

$$a_i(t)=\sum_{m=1}^{R}\alpha_{i,m}\pi_m(t),$$

$$b_l(t)=\sum_{m=1}^{R}\beta_{l,m}\pi_m(t) \tag{2.6}$$

where $\alpha_{i,m}$ and $\beta_{l,m}$ represent the expansion parameters, R is the maximum number of basis sequences, $\pi_m(t)$, $m=1,2,\cdots,R$ are a set of basis functions. Substituting (2.6) into (2.5) yields,

$$y(t)=\sum_{i=1}^{p}\sum_{m=1}^{R}\alpha_{i,m}\pi_m(t)y(t-i)+\sum_{l=1}^{q}\sum_{m=1}^{R}\beta_{l,m}\pi_m(t)u(t-l)+e(t), \tag{2.7}$$

Once proper basis functions have been chosen, new variables can be defined such that

$$y_m(t-i)=\pi_m(t)y(t-i),\quad u_m(t-l)=\pi_m(t)u(t-l). \tag{2.8}$$

Substituting (2.8) into (2.7), yields,

$$y(t)=\sum_{i=1}^{p}\sum_{m=1}^{R}\alpha_{i,m}y_m(t-i)+\sum_{l=1}^{q}\sum_{m=1}^{R}\beta_{l,m}u_m(t-l)+e(t), \tag{2.9}$$

The model in (2.9) can be written down in the following form:

$$y(t)=\varphi^T(t)\theta(t)+e(t) \tag{2.10}$$

where

$$\varphi(t)=\left[y_m(t-1),\cdots,y_m(t-p),u_m(t-1),\cdots,u_m(t-q)\right]^T \tag{2.11}$$

is the regression vector and

$$\theta(t)=\left[\alpha_{1,m},\cdots,\alpha_{p,m},\beta_{1,m},\cdots,\beta_{q,m}\right]^T \tag{2.12}$$

is the model coefficient vector, and the upper script $'T'$ indicates the transpose of a vector or a matrix. Note that all the entries $\alpha_{i,m}$ and $\beta_{l,m}$ in the coefficient

vector are now constants. Thus, the original time-varying AR model (2.4) and ARX model (2.5) have now been converted into a time invariant model (2.10).

Equation (2.10) is a standard linear regression model that can be solved by using a linear least squares algorithm. Let $\hat{a}_i(t)$, $\hat{b}_l(t)$ be the estimate of $a_i(t)$, $b_l(t)$ and $\hat{\sigma}_e^2$ be the estimate of σ_e^2, respectively. The time-dependent spectral function associated to the TVAR model (2.4) and TVARX model (2.5) are defined as, respectively,

$$H(f,t) = \frac{\hat{\sigma}_e^2}{\left|1 - \sum_{i=1}^{p} a_i(t) e^{-j2\pi i f / f_s}\right|^2} \quad ,$$

and

$$H(f,t) = \left| \frac{\sum_{l=0}^{q} \hat{b}_l(t) e^{-j2\pi l f / f_s}}{1 - \sum_{i=1}^{p} \hat{a}_i(t) e^{-j2\pi i f / f_s}} \right|^2 , \tag{2.13}$$

where $j = \sqrt{-1}$, i.e. j is the square root of -1, and f_s is the sampling frequency. Note that the spectral function (2.13) is continuous with respect to the frequency f and thus can be used to produce spectral estimates at any desired frequencies up to the Nyquist frequency $f_s/2$. However, the frequency resolution is primarily not infinite, but is determined by the underlying model order and the associated parameters.

2.3.4 Multi-wavelet basis functions

When a time-varying system is subject to rare but abrupt jumping, the estimated parameters from conventional adaptive algorithms cannot track the variations of the true system parameters in the vicinity of these jumping locations, resulting in the so

called “lag” in the estimation. The common parametric approach, for establishing the time-varying model based on the basis function expansion approach, which shows excellent capability on tracking coefficients changing over time, can be used to mitigate the effect of “lag” estimation. Various basis functions including the Fourier series (Pachori & Sircar 2008), Legendre polynomial (Niedzwiecki 1988), Walsh function (Tsatsanis & Giannakis 1993; Zou et al. 2003), Haar wavelet and multi-wavelet (Chen et al. 2006; Ghanem & Romeo 2000; Tsatsanis & Giannakis 1993; Wei & Billings 2002; Wei & Billings 2004; Wei & Billings 2006a; Wei & Billings 2006b; Wei et al. 2004a), have been applied to describe time-varying linear and nonlinear model coefficients. Selecting the proper basis functions is a key to the success of this approach. In practice, each family of basis functions possess its own unique tractability and accuracy. Zou et al. (2003) has investigated that the Legendre polynomial performed well for the coefficients that change smoothly with time, the Walsh and Haar functions, however, were good for piecewise stationary TV coefficients that have sharp variations or piecewise changes. Wei and Billings (2002) has introduced a TV modelling approach approximated by a finite number of multi-wavelet functions which outperforms many other approximation schemes and the results have shown the effectiveness and applicability of this method.

2.3.5 The orthogonal least squares algorithm

In system identification and modelling, especially in nonlinear system identification, modelling and signal processing, the orthogonal least square regression (OLS) algorithm (Billings et al. 1989; Chen et al. 1989; Wei & Billings 2004; Wei & Billings 2006a; Wei & Billings 2006b; Wei et al. 2006; Zhu & Billings 1996) has been proved to be a very effective algorithm to deal with multiple dynamical regression problems, which involve a great number of candidate model terms or regressors that may be highly correlated, and determine significant model terms or the model structure and the associated parameter

estimates. The OLS algorithm involves a stepwise orthogonalization of the regressors and a forward selection of the relevant terms based on the error reduction ratio (ERR) criterion (Billings et al. 1989; Chen et al. 1989). Recently, the OLS algorithm and its variants have been successfully applied in a variety of fields including system identification and modelling.

The most difficult problem in linear and nonlinear system identification and modelling is determining the structure or architecture of the model based on a limited set of observed or measured data. Assuming that no a priori knowledge of the form of the nonlinear functional is available, a practical solution is to approximate the system from the available data using a known set of basis functions or regressors that belong to a given function class. Typical regressor classes used in linear and nonlinear system identification and modelling include polynomial and rational functions, radial basis functions (RBF) and wavelet neural networks.

The classical OLS regression algorithm includes the following steps:

1) Orthogonalize the regressors to remove the correlations between these variable;
2) Select significant terms using the ERR criterion;
3) Estimate the corresponding parameters for the selected terms;
4) Validate Model.

Model validation is an essential and important step in system identification and modelling. Model validation for linear systems is well established. The residuals should be an unpredictable sequence if the model structure is correct and the estimated parameters are unbiased. Hence the auto-correlation function of the residuals and the cross-correlation function between the residuals and the input have been widely used in linear model validation. Model validation methods based

on higher order correlation test were introduced for nonlinear systems (Billings & Voon 1986; Billings & Zhu 1994). Model validation of systems identified determines and confirms whether the model can adequately describe the underlying dynamics. Several methods are used to test the model validity of the system identified. The most used method is judging the quality of the one step ahead prediction (OSA) errors or the model predicted output (MPO). An alternative method is to compare specific dynamical characteristics of real systems and the model, for example, largest Lyapunov exponent, bifurcation dimension, correlation dimension and the wavelet entropy etc.

The application of the common model structure selection approach that involves model structure selection and model parameter estimation based on the OLS algorithm is also investigated in Chapter 5. Once a common model structure has been obtained, an online sliding-windowing recursive least squares (SWRLS) algorithm can then be applied to estimate the time-varying model parameters. The time varying common structure (TVCS) model can then be obtained and applied to dynamically track and capture the transient variation of the nonstationary signals. The central purpose of obtaining the TVCS model is focused on nonlinear time-varying parametric modelling in the frequency domain. For example, the time-varying parameter results estimated from the TVCS model cannot only provide the transient local information of the biomedical signals, but can also be applied in nonlinear time-dependent parametric spectral analysis in the frequency domain to extract more features from the electrophysiological signals.

As to the issue of the model order determination, this can be solved by using some model order determination criteria such as the well-known Akaike information criterion (AIC) (Akaike 1974), Bayesian information criterion (BIC) (Schwarz 1978)(Schwarz 1978) and the modified generalized cross-validation (GCV) criteria (Billings & Wei 2007).

2.4 Frequency-domain analysis of nonlinear systems

A NARMAX model in the frequency domain can be computed by the nonlinear frequency response functions also called the generalised frequency response functions (GFRFs). The traditional representation of nonlinear systems based on the Volterra series can be defined by

$$y(t)=\sum_{p=1}^{\infty} y_p(t), \tag{2.14}$$

where

$$y_p(t)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} h_p(\tau_1,\cdots,\tau_p)\prod_{i=1}^{p} u(t-\tau_i)\,d\tau_i\,, \tag{2.15}$$

Taking the multiple Fourier transform of the p th order Volterra kernel yields the p th order GFRFs function expression

$$H_1(f_1,\cdots,f_p)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} h_p(\tau_1,\cdots,\tau_p)e^{-j2\pi(f_1\tau_1+\cdots+f_p\tau_p)}d\tau_1\cdots d\tau_p\,, \tag{2.16}$$

We can obtain the nonlinear p th order impulse response formula with the inverse Fourier transform

$$h_p(\tau_1,\cdots,\tau_p)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} H_p(f_1,\cdots,f_p)e^{j2\pi(f_1\tau_1+\cdots+f_p\tau_p)}df_1\cdots df_p \tag{2.17}$$

Substituting equation (2.17) into equation (2.15) and carrying out the multiple integrals on $\tau_1,\cdots,\tau_p$, we can then obtain

$$y_p(t)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} H_p(f_1,\cdots,f_p)\prod_{i=1}^{p} U(f_i)e^{j2\pi f_i t}df_i\,, \tag{2.18}$$

where $U(f)$ indicates the input spectrum.

Billings and Jones (1990) proposed the probing method to calculate the GFRFs. For example, Let the input $u(t)$ be a sum of K sinusoids

$$u(t)=\sum_{k=1}^{K}A_k e^{j2\pi f_k t}, \tag{2-19}$$

where A_k means the amplitudes and f_k is any real number. According to equation (2.15), the p th order output can be shown as follows

$$y_p(t)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}h_p(\tau_1,\cdots,\tau_p)\prod_{i=1}^{p}\sum_{k=1}^{K}A_k e^{j2\pi f_k(t-\tau_i)}d\tau_i$$

$$=\sum_{k=1}^{K}\cdots\sum_{k_l}^{K}\prod_{i=1}^{l}A_k e^{j2\pi f_{ki}(t-\tau_i)}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}h_k(\tau_1,\cdots,\tau_k)\prod_{i=1}^{p}A_p e^{j2\pi f_{ki}(t-\tau_i)}d\tau_i, \tag{2.20}$$

Substituting impulsive response equation (2.17) into equation (2.20) yields

$$y_p(t)=\sum_{k=1}^{K}\cdots\sum_{k_p}^{K}\left[A_{k1},\cdots,A_{k_p}H_p\left(f_{k_1},\cdots,f_{k_p}\right)\right]e^{-j2\pi\left(f_{k_1}+\cdots+f_{k_p}\right)t}, \tag{2.21}$$

It should be worth stressing that the Fourier transform of equation (2.21) is a sum of delta functions. Setting $K=p$ and amplitude $A_k=1$ for all $k=1,\cdots,p$, for equation (2.21), we can then obtain

$$y_p(t)=\sum_{k=1}^{p}\cdots\sum_{k_p}^{p}H_p\left(f_{k_1},\cdots,f_{k_p}\right)e^{-j2\pi\left(f_{k_1}+\cdots+f_{k_p}\right)t}, \tag{2.22}$$

When the system input is given by equation (2.19) with $K=p$ and $A_k=1$, then the GFRFs can be obtained by equating the coefficients of $p!\left\{j2\pi\left(f_1+\cdots+f_p\right)\right\}$ in the system output.

2.5 EEG Granger Causality analysis

Mathematical measures including coherence, correlation, phase synchronization or mutual information are usually applied to evaluate and describe the interactions between groups of neurons to investigate neural connections. However, in neurobiology, the analytical results from these methods cannot distinguish directions of flow between two cortical sites or causality. To assess the directionality of neuronal interactions and understand the cooperative nature of

neural computation, Granger causality is a fundamental tool for the description of causal interaction of two time series. Especially, if the prediction error of the first time series at the present time is reduced by including past measures of the second time series in the linear regressor model, then the second time series is said to have a causal influence on the first one. By exchanging the role of the two time series, we can address the question of causal influence in the opposite direction. From the definition of causality, the flow of time plays a significant role in allowing inferences from the time series. It should be noted that Granger causality was formulated for the linear model case, its application to nonlinear systems may not be appropriate. Recently some attempts have been proposed to extend the linear Granger causality to the nonlinear case. For example, Ancona et al. (2004) has investigated that a radial basis function (RBF) method has been applied to model data to evaluate the causality. Gourevitch et al. (2006) have evaluated the measures of Granger causality on some linear and nonlinear models, and they have also investigated some of the properties and drawback for linear and nonlinear Granger causality.

All the previous methods are based on the time-invariant linear and nonlinear models to evaluate the Granger causality. Some evolution of nonlinear time series models has shown that standard linear vector autoregression models cannot capture the dynamics behaviour of many electrophysiological data adequately. However, a lot of components of neural systems are not linear and are nonlinear at a stationary level. Ding et al. (2000) used a short-time windows technique, which only requires the stationarity of the signal within short-time windows, and enables the construction of a time-varying Granger causality. Hesse et al. (2003) studied and focused on the recursive TV estimation of the Granger causality. The proposed approach is based on the adaptive recursive fit of the VAR model with TV parameters by means of a generalised RLS algorithm, and overcomes the requirement of stationarity of the signals and thus captures the observation of transient directed causal networks. It is worth stressing that the most important

property of the Granger causality is its positivity. However, negative values of Granger causality appeared in (Hesse et al. 2003). The main reasons of occurrence of negative values include the following. Firstly, the RLS algorithm applied to estimate the TV parameters in the VAR models results in a tracking of the autoregressive parameters around their theoretical values. Negative values of estimated Granger causality are possible due to the non-consistency of the parametric estimations. Second, the order in the VAR model investigated is fixed. This order may be not optimal for single EEG channel pairs. Third, the causal method discussed is only based on TV linear VAR models and considers transient linear causality. However, a lot of components of neural systems such as electrophysiological data are not linear. Some authors have showed evidence of nonlinearities in EEG time series (Palus 1996).

To solve the issues aforementioned, in Chapter 6, a novel linear and nonlinear TV parametric modelling approach on the basis of data-driven methods is investigated for the adaptive estimation of nonstationary EEG Granger causality processes to detect the transient dynamic causal directional interactions between EEG signals within time intervals. The approach proposed allows identification of the directed information flow between brains areas. Two numerical examples, one for an artificial signal, where the exact answers of causal influences are known, and another for EEG signals are given to show the effectiveness and applicability of the proposed TV linear and nonlinear Granger causality method between EEG single-channel pairs. The more detail information such as methods and result discussions can be discussed in Chapter 6.

2.6 Nonlinearity detection for time series

To illustrate the necessity to evaluate the nature of a real-world signal comprising both linear and nonlinear components prior to choosing a mathematically tractable

model, recently, a great number of different measure approaches have been investigated to detect nonlinearity in time series such as approximate entropy, correlation dimension, largest Lyapunov exponents, higher order statistics and nonlinear prediction error (NPE) (Farmer & Sidorowich 1987; Sugihara & May 1990). However, some of the measure approaches cannot distinguish chaos from coloured noise due to the low discrimination power or the correlation dimension, or requirement of long time series that is generally not available in the real world. Particularly, the NPE approach based on phase-space reconstruction can give either better or comparable performance than other approaches (Schreiber & Schmitz 1997).

The surrogate data method that belongs to the family of statistical tests known as hypothesis testing can also provide a rigorous statistical approach to nonlinear feature detection of a time series. The approach of surrogate data has been extensively used in the context of statistical nonlinearity testing underlying experimental data and is widely exploited for evaluating the capability of nonlinear indexes to test the presence of nonlinearity providing that the observed time series is stationary. Due to the assumption of stationarity, the surrogate time series approach is a time-invariant method that has statistical distribution properties which are invariant to translation of the original time series. However, stationary time sequences to be analyzed in electrophysiological data are often impossible to find even in short-term recordings. With regard to the approach of surrogate data, ambiguities between nonlinearity and nonstationarity might arise when the null hypothesis of a time-invariant linear process is rejected with a nonlinear discriminating statistics.

To provide a unifying method for detecting the nature of electrophysiological experimental data in real-world signals, Mandic et al. (2008) have investigated a new delay vector variance (DVV) approach for characterizing a time series, and have also assessed the performance in the presence of nonlinearity detection. The

case studies discussed illustrated the effectiveness of the proposed method for the qualitative assessment of functional magnetic resonance imaging (fMRI) data. In order to detect the presence of nonlinear dynamics of nonstationary time series potentially, Fase et al. (2009) studied an approach based on generating TV surrogated data according to the null hypothesis of TV linear stochastic process. The proposed method is that first fitting a TV autoregressive (AR) model to the original time series, and then regressing the model coefficients estimated with random replacements of the model residuals to generate the TV surrogated time series. Model identification procedure is expanded the TV model coefficients onto a finite set of pre-defined basis functions, the classical nonlinear index criteria, for example, iterative amplitude adjusted Fourier transform (*i*AAFT) method that produces surrogates with identical signal distributions and approximately identical amplitude spectra as the original time series (Schreiber & Schmitz 1996), is applied as a discriminating statistic. The analysis results have been shown that using TIV surrogate, linear nonstationary time series may be regarded as nonlinear erroneously and weak TV nonlinearities may still not be revealed, while the application of TVAR surrogates increases the probability of a correct interpretation markedly. In the following of our research work, we can also apply the TVAR method with the basis function expansion in our research group to detect the nonlinearity of time series, and nonlinearity analysis results should increase the probability of a more accurate interpretation with respect to the definition of linearity that has been adopted.

2.7 Summary

In real-world applications, most biomedical signals are nonstationary and often involve numerous TV and transient components associated with underlying psychological or physiological activities. Therefore, the time-varying NARMAX polynomial model scheme that is simple and effective is one of the commonly

popular approaches to describe the dynamics of nonstationary signals or capture the transient variation of such signals in both time and frequency domains. Time-varying model coefficients can be applied to estimate time-frequency distribution of biomedical signals. Generally, identification and estimation of a time-varying model can commonly be classified into two categories, namely: 1) classical adaptive recursive algorithm and Kalman filtering and 2) basis function expansion scheme that uses pre-defined deterministic basis expansion to model the coefficient variations where the TV coefficients are approximated by a linear or nonlinear combination of known basis functions. The recursive algorithms or basis functions expansion approach is highly dependent on the unknown parameters and the performance will be degraded when parameters are estimated inappropriately.

In this chapter, the adaptive recursive algorithms and its performance comparison have been investigated and basis function expansion method with OLS algorithm has also been discussed to estimate the TV model coefficients. The identification model is then applied to various applications using time-frequency analysis techniques that have been systematically and extensively applied to the study of biomedical signals. Different time-frequency distribution approaches for linear and nonlinear cases have been discussed. The time-frequency analysis of the parametric modelling method has a more accurate time-frequency resolution than conventional time-frequency distribution analysis methods like HHT, STFT and CWT.

Granger causality for linear and nonlinear models has been developed in recent years. Some model and evaluated methods have even been applied to assess the directed causal influence in electrophysiological signals. However, there are virtually no results on identification and modelling of TV nonlinear systems to analyse the Granger causality influence between biomedical channel pairs. The different analytical approaches of nonlinearity detection for time series have been introduced briefly. All these topics will be discussed in detail in the following chapters.

Chapter 3

Identification of Time-Varying Systems Using Multi-wavelet Basis Functions

3.1 Introduction

Many processes are inherently TV and cannot effectively be characterised using time invariant models. Modelling and analysis of TV systems is often a challenging problem. One feature of TV systems is that such signals contain nonstationary transient events. One approach to characterise such nonstationary processes is to employ TV parametric models for example the TV autoregressive with exogenous input (TVARX) model (Peng et al. 2007), or simply the TVAR model (Chowdhury et al. 2006). Two main classes of methods can be used to resolve the TVARX and TVAR model estimation problem. The first popular approaches including traditional RLS algorithm (Ljung & Gunnarsson 1990), the Kalman filter and the Random Walk Kalman Filter (RWKF) algorithms (Chowdhury 2000; Morbidi et al. 2008; Niedzwiecki 1994) have been derived and applied to analyse time varying dynamics of linear systems, and the second class of approaches treat the evolution of the time-varying coefficients to be linear or nonlinear combinations of a number of basis functions with some good properties; this allows the associated TV coefficients to be expanded as a finite sequence of pre-determined basis functions for example wavelet basis functions (Chon et al. 2005; Wei & Billings 2002; Wei et al. 2010).

An attractive approach is to expand the TV coefficients using wavelets as the basis functions. Wavelets have been proved to be a valuable tool for signal processing and have been shown to possess excellent linear or non-linear approximation

properties which outperform many other approximation schemes and are well suited for approximating general nonstationary signals, even those with very sharp or abrupt discontinuities. Wavelets have also successfully been used in system identification and modelling (Billings & Wei 2005a; Chang & Liu 2006; Chowdhury & Aravena 1998; Tsatsanis & Giannakis 1993).

In this Chapter a new wavelet multi-resolution parametric modelling and identification technique for the identification of systems with TV parameters is proposed, where the associated time dependent parameters are approximated using a set of multi-wavelet basis functions, which transforms the TV identification problem into a time-invariant parametric expansion. The identification of the model parameters can then be achieved by adopting a block LMS algorithm. One advantage of the proposed approach, which combines wavelet approximation theory with a block LMS algorithm, is that the new wavelet based algorithm can be used to track very rapidly or even sharply varying processes. The novel approach proposed can thus track rapid time variation and is more suitable for the estimation of process parameters of inherently nonstationary processes. A multi-wavelet basis function approach is used because of the ability to capture the signals characteristics at different scales. Two examples, one for a synthetic data set and a second for a mechanical system are given to illustrate the capability and efficacy of the proposed method. It is shown that the proposed method can produce much better tracking performance compared with traditional LMS and RLS approaches.

3.2 Methodology

Consider an input-output relationship of a TVARX process which is described by the following equation:

$$y(t)=\sum_{i=1}^{p} a_i(t)y(t-i)+\sum_{l=1}^{q} b_l(t)u(t-l)+e(t), \tag{3.1}$$

where u, y and $e(t)$ are the sampled measurable input, output and prediction error signals, $a_i(t)$ and $b_l(t)$ are the TV parameters to be determined, p and q are the maximum model orders, and t represents discrete time. The proposed method expands the TV parameters $a_i(t)$ and $b_l(t)$ onto multi-wavelet basis functions $\pi_m(t)$ for $m=1,2,\cdots,R,$ such that the following expressions hold:

$$a_i(t)=\sum_{m=1}^{R}\alpha_{i,m}\pi_m(t), \qquad b_l(t)=\sum_{m=1}^{R}\beta_{l,m}\pi_m(t), \tag{3.2}$$

where $\pi_m(t),\ m=1,2,\cdots,R$ are wavelet basis functions, $\alpha_{i,m}$ and $\beta_{l,m}$ are the expansion parameters, R is the maximum number of basis sequences. Substituting (3.2) into (3.1), yields (3.3),

$$y(t)=\sum_{i=1}^{p}\sum_{m=1}^{R}\alpha_{i,m}\pi_m(t)y(t-i)+\sum_{l=1}^{q}\sum_{m=1}^{R}\beta_{l,m}\pi_m(t)u(t-l)+e(t), \tag{3.3}$$

Once basis functions have been chosen (see next section), new variables can be defined as below

$$y_p(t-i)=\pi_m(t)\,y(t-i), \qquad u_q(t-l)=\pi_m(t)u(t-l). \tag{3.4}$$

Substituting (3.4) into (3.3) yields,

$$y(t)=\sum_{i=1}^{p}\sum_{m=1}^{R}\alpha_{i,m}y_p(t-i)+\sum_{l=1}^{q}\sum_{m=1}^{R}\beta_{l,m}u_q(t-l)+e(t), \tag{3.5}$$

Model (3.5) can be rewritten as

$$y(t)=\varphi^T(t)\theta(t)+e(t), \tag{3.6}$$

where

$$\varphi(t)=\left[y_p(t-1),\cdots,y_p(t-p),u_q(t-1),\cdots,u_q(t-q)\right]^T, \tag{3.7}$$

and

$$\theta(t)=\left[\alpha_{1,m},\cdots,\alpha_{p,m},\beta_{1,m},\cdots,\beta_{q,m}\right]^T, \tag{3.8}$$

$\varphi(t)$ and $\theta(t)$ are the regression vector and coefficient vector, respectively, and the upper script $'T'$ indicates the transpose of a vector or a matrix. Eq. (3.5) and

Eq. (3.6) show that the TVARX(p,q) model can now be treated as a time invariant model, since $\alpha_{i,m}$ and $\beta_{l,m}$ are now time-invariant.

3.3 The Multi-Wavelet Basis Functions

From wavelet theory (Chui 1992; Mallat 1989), a signal can be decomposed into components spanned by the scaling and shifting wavelet basis functions at different resolutions. As defined by (Mallat 1999), any finite energy function $f(x)$, *i.e.*, functions for which the integral $\int_{-\infty}^{\infty}|f(x)|^2 dx$ exists with a finite value, can be arbitrarily approximated by

$$f(x)=\sum_{k=-\infty}^{\infty}\alpha_{j_0 k}\phi_{j_0}(x)+\sum_{j=j_0}^{\infty}\sum_{k=-\infty}^{\infty}\beta_{j,k}\psi_{j,k}(x), \tag{3.9}$$

where $\phi_{j_0,k}(x)=2^{j_0/2}\phi\left(2^{j_0}x-k\right)$ and $\psi_{j,k}(x)=2^{j/2}\psi\left(2^{j}x-k\right)$ are the scaled and translated version of the scaling function $\phi(x)$ and the mother wavelet $\psi(x)$, $\alpha_{j_0,k}$ and $\beta_{j,k}$ are the wavelet decomposition coefficients at the scale level j_0 and j, respectively. The first summation indicates a low resolution or coarse approximation of $f(x)$ at the scale level j_0 in Eq. (3.9). For each j, a finer or higher resolution function including more detail of $f(x)$ is added in the second summation. Like in any regression representation of functions, $\phi_{j_0,k}(x)$ and $\psi_{j,k}(x)$ can be interpreted as regressions. Eq. (3.9) can be reduced to

$$f(x)=A_{j_0}+\sum_{j=j_0}^{J}D_j\ , \tag{3.10}$$

where $A_{j_0}=\sum_{k=-\infty}^{\infty}c_{j_0,k}\phi_{j_0,k}$ is the approximation at level j_0 and $D_j=\sum_{k=-\infty}^{\infty}d_{j,k}\psi_{j,k}$ is the detail at level j. When the higher resolution components make little

contribution to $f(x)$, the function can be approximated by a truncated multi-resolution wavelet expansion from j_0 up to J $(j_0 \le J)$ as given in Eq. (3.10). It is worth stressing that the multi-resolution wavelet of a signal is not unique and depends on the type of mother wavelet and scaling function that are used. In (Ahuja et al. 2005), properties of various types of wavelets including orthogonality, support, regularity and time-frequency localization window were discussed for various wavelet types. It was concluded that B-spline mother wavelets could be justified for wavelet-based high resolution image sequences. Cardinal B-splines are an important class of basis functions that can form multi-resolution wavelet decompositions (Chui 1992). The first order cardinal B-spline is the well-known Haar function defined as

$$B_1(x) = \chi_{[0,1)} = \begin{cases} 1, & x \in [0,1), \\ 0, & otherwise. \end{cases} \tag{3.11}$$

The second, third and fourth order cardinal B-splines $B_2(x)$, $B_3(x)$ and $B_4(x)$ are given in (Wei & Billings 2006a). Note that the polynomial degree of *m*th order B-spline is equal to (m-1). For detailed discussions on cardinal B-splines, the associated wavelets and details of the excellent approximation properties, see (Chui 1992).

One attractive feature of cardinal B-splines is that these functions are completely supported, and this property enables the operation of the multi-resolution decomposition (3.9) to be much more convenient. For example, the m th order B-spline is defined on $[0, m]$, with the scale and shift indices j and k, for the family of the functions

$$\phi_{j,k}(x) = 2^{j/2} B_m\left(2^j x - k\right), \quad 0 \le 2^j x - k \le m. \tag{3.12}$$

Assume that the function $f(x)$ that is to be approximated with decomposition (3.9) is defined within $[0,1]$, then for any given scale index (resolution level) j,

based on $x \in [0,1]$ and $0 \le 2^j x - k \le m$, the effective values for the shift index k, are restricted to the collection $\{k: -m < k \le 2^j\}$ with the $B_m(m) = 0$.

Note that while the first and second order B-splines $B_1(x)$ and $B_2(x)$ are non-smooth piecewise functions, which would perform well for signals with sharp transients and burst-like spikes, B-splines of higher order would work well on smoothly changing signals. Motivated by this consideration, this study proposes using multi-wavelet basis functions for TVARX model identification. Two examples of the new multi-wavelet based algorithm are given in the following.

Take the B-splines of order from 1 to 4 as an example, and consider the decomposition (3.10). Let $\Gamma_m = \{k: m < k \le 2^j\}$, $m = 1,2,3,4$ and $\phi_k^{(m)}(x) = 2^{j/2} B_m(2^j x - k)$, $k \in \Gamma_m$. The TV coefficients $a_i(t)$ and $b_l(t)$ in model (3.1) can then be approximated using a combination of functions from the families $\{\phi_k^{(m)}: m = 1,\cdots,4;\ k \in \Gamma_m\}$. Several criteria for the determination of j were given in (Wei & Billings 2002). For the B-splines wavelet used in this Chapter, the lower level j_0 and higher level j in Eq. (3.10) can be chosen $j_0 = j = 3$. However, the higher level j can be chosen as larger than j_0 as possible provided that the computation effort is not a concern. Therefore, one such combination can be represented as,

$$\begin{aligned} a_i(t) &= \sum_{k\in\Gamma_\eta} c_{i,k}^{(\eta)} \phi_k^{(\eta)}\left(\frac{t}{N}\right) + \sum_{k\in\Gamma_r} c_{i,k}^{(r)} \phi_k^{(r)}\left(\frac{t}{N}\right) + \sum_{k\in\Gamma_r} c_{i,k}^{(s)} \phi_k^{(s)}\left(\frac{t}{N}\right) \\ b_l(t) &= \sum_{k\in\Gamma_\eta} d_{j,k}^{(\eta)} \phi_k^{(\eta)}\left(\frac{t}{N}\right) + \sum_{k\in\Gamma_r} d_{j,k}^{(r)} \phi_k^{(r)}\left(\frac{t}{N}\right) + \sum_{k\in\Gamma_r} d_{j,k}^{(s)} \phi_k^{(s)}\left(\frac{t}{N}\right) \end{aligned} \tag{3.13}$$

where $1 \le \eta < r < s \le 4$, $t = 1,2,\cdots,N$, and N is number of observations of the signal. Modelling experiences and simulation results with a large number of

experiments have shown that for most TV problems, the choice of $\eta = 2,\ \ r = 3$ and $s = 4$ works well to recover the TV coefficients. Note that when choosing B-splines it should try to include different types of polynomials, for example, the second order B-spline performs well for time varying coefficients that have sharp variations or piecewise changes, while the third and fourth order B-splines can work well for most smoothly and slowly varying coefficients, to track both fast and slowly varying trends in this Chapter. If, however, there is strong evidence that the time-dependent coefficients have sharp changes, then the inclusion of the first and second order B-splines would work well. The decomposition (3.13) can easily be converted into the form of (3.2), where the collection $\{\pi_m(t):\ m = 1, 2, \cdots, R\}$ is replaced by the union of the three families: $\{\phi_k^{(\eta)}(t):\ k \in \Gamma_\eta\}$, $\{\phi_k^{(r)}(t):\ k \in \Gamma_r\}$ and $\{\phi_k^{(s)}(t): k \in \Gamma_s\}$. Further derivation can then lead to the standard linear regression equation (3.5). Eq. (3.13) and Eq. (3.5) indicate that the initial full regression equation (3.5) may involve a great number of free time-dependent TV parameters, and least squares type algorithms may fail to produce reliable results for such ill-posed problems. These problems, however, can easily be overcome by performing an effective online block LMS algorithm, the resulting recursive coefficient estimates $c_{i,k}$ and $d_{l,k}$ in Eq. (3.13) will then be used to recover the TV coefficients $a_i(t)$ and $b_l(t)$ in the TVARX in model (3.1).

3.4 Model Identification and Parameter Estimation with A Block Least Mean Square Approach

The conventional LMS algorithm and normalized LMS (Haykin 2003; Shynk 1992) have been proven to be very effective to deal with dynamic regression problems. However, the performance of these algorithms is sensitive to the selection of step

sizes and additional noise. In this Chapter we introduce a block LMS algorithm to estimate the TV parameters in the model. Table 3.1 presents a summary of the block LMS algorithm, where the step size μ is divided by the maximum eigenvalue of the correlation matrix C for a potentially-faster algorithm convergence time. An important issue that needs to be considered in the design of a block adaptive filter is how to choose the block size L. From Table 3.1 we observe that the operation of the block LMS algorithm holds true for any integer value of L equal to or greater than unity. Nevertheless, the option of choosing the block size L equal to the filter length (that is, the number of TV parameter coefficients in model (3.1)) M is referred in most applications as block adaptive filtering. This choice has been justified by Clark et al. (1981) based on the following observations: When $L>M$, redundant operations are involved in the adaptive process, because then the estimation of the gradient vector (computed over L points) uses more input information than the filter itself. When $L<M$, some of the tap weights in the filter are wasted, because the sequence of tap inputs is not long enough to feed the whole filter.

It thus appears that the most practical choice is $L=M$. For $L=1$, the block LMS algorithm simplifies to the normalized LMS algorithm, where C is a scalar. For $L>1$, Table 3.1 summarizes the block LMS algorithm, where C is a square matrix.

The block LMS approach leads to two significant advantages over the conventional LMS algorithm: 1) for $L=1$, potentially-faster convergence speeds for both correlated and whitened input data (Clark et al. 1981), and stable behaviour for a known range of parameter values ($0<\mu<2/\lambda_{\max}$, stability condition) independent of the input data correlation statistics (Goodwin & Sin 1984; Nagumo & Noda 1967); 2) for $L>1$, block processing of data samples, a block of samples of the filter input and desired output are collected and then processed together to obtain a

block of output samples. A good measure of computational complexity in a block processing system is given by the number of operations required to process one block of data divided by the block length. An implementation of the block LMS algorithm is more computationally efficient.

Table 3.1 Summary of the block least mean square algorithm

Definition

$\{u(1),u(2),\cdots\}$, input signal samples

$\{y(1),y(2),\cdots\}$, desired signal samples correlated with input signal samples

L block size

M, filter length (M is equal to the sum of p and q in model (1)),

μ, step-size,

a, a small positive constant,

$\lambda_{\max}=\max\{eig(C)\}$, maximum eigenvalue of the correlation matrix

$C=E\left(X^T(k)X(k)\right)$,

$W(k)=\left[w_1(k),\cdots,w_M(k)\right]^T$, a vector of weights.

Initial Conditions:

$\hat{W}(0)=0_{M\times 1}$,

Computation: at the k th iteration, for each new block of M input samples, compute

$$X(k)=\begin{bmatrix} u(kM) & \cdots & u\left((k-1)M+1\right) \\ \vdots & \ddots & \vdots \\ u\left((k+1)M-1\right) & \cdots & u(kM) \end{bmatrix},$$

$d(k)=\left[d(kM),\cdots,d\left((k+1)M-1\right)\right]^T$,

$y(k)=X(k)\hat{W}(k)$,

$e(k)=d(k)-y(k)$,

$\phi(k)=X^T(k)e(k)$,

$$\hat{W}(k+1)=\hat{W}(k)+\frac{\mu}{(a+\lambda_{\max})}\phi(k).$$

Dimensions:

$\hat{W}(k)$ $M\times 1$; $X(k)$ $M\times M$; $d(k)$ $M\times 1$; $y(k)$ $M\times 1$; a 1×1;

$e(k)$ $M\times 1$; $\phi(k)$ $M\times 1$; μ 1×1; C $M\times M$; $\lambda_{\max}$ 1×1.

In the present study, the block LMS algorithm described above is applied to refine and solve the regression equation (3.5). This includes a model identification and time-dependent TV parameter estimation. The resultant estimates will then be used to recover the TV coefficients $a_i(t)$ and $b_l(t)$ in the TVARX model (3.1).

To determine the proper model size selection for the TVARX case given by Equation (3.5), we exploited the fact that time-dependent or TV coefficients are expanded onto multi-wavelet basis functions to give a time invariant system. Hence, the modified generalized cross-validation (GCV) criteria (Billings et al. 2007) can be employed to account for the increased number of time invariant coefficients for the system (3.5). This modified criterion allowed us not only to select the proper model size, but also to choose an appropriate number of multi-wavelet basis functions onto which the TV coefficients are expanded.

3.5 Simulation Examples

To verify the performance of the multi-wavelet basis functions approach, two examples will be studied. The first is a simulated experiment with measurement SNR's (Signal to Noise Ratio) presented. Consider the following TVARX$(1,1)$ model,

$$y(t)=a_1(t)y(t-1)+b_1(t)u(t-1)+e(t), \tag{3.14}$$

The process parameters $a_1(t)$ and $b_1(t)$ will be varied in different ways and the output $y(t)$ is observed for the system input $u(t)$ which was a Pseudo-Random Binary Sequence (PRBS) (Leontaritis & Billings 1987) where $e(t)$ is a discrete white noise sequence. The system parameters are estimated using the normalized LMS approach, RLS approach and the block LMS approaches based on multi-wavelet basis functions, respectively, so that these three methods can be

compared.

The TV parameter variations were designed to change in an abruptly varying manner as

$$a_1(t)=\begin{cases}-0.1 & 0\le t\le 0.3\\ 0.9 & 0.3<t\le 0.5\\ -0.5 & 0.5<t\le 0.7\\ 0.6 & 0.7<t\le 1.\end{cases},\qquad b_1(t)=\begin{cases}0.1 & 0\le t\le 0.2\\ -0.5 & 0.2<t\le 0.4\\ 0.8 & 0.4<t\le 0.7\\ 0.3 & 0.7<t\le 1\end{cases}, \tag{3.15}$$

The PRBS input signal $u(t)$ is a frequency rich signal. The input signal is of 1 second duration and the sampling frequency was 1000 Hz. The output is a nonstationary signal with a SNR = 19.4024dB. Figure 3.1 shows the true and estimated values of parameters $a_1(t)$ and $b_1(t)$ respectively for the noise measurement of SNR = 19.4024dB using the normalized LMS algorithm with $L=1,\ \mu=0.6$. The estimated parameters follow the true parameter variations quite well. Figure 3.2 shows the results of the true and estimated values for parameters $a_1(t)$ and $b_1(t)$ respectively with the same noise measurement of 19.4024 dB using the RLS algorithm with forgetting factor $\mu=0.90$. The RLS approach obtains smoother estimates but does not faithfully track the rapidly changing TV parameters. Figure 3.3 shows the results of true and estimated values for parameters $a_1(t)$ and $b_1(t)$ respectively with the noise measurement of 19.4024 dB using the new block LMS based algorithm with $L=2,\ \mu=1$ using multi-wavelet basis functions. The estimated parameters follow the true parameters variations extremely well picking up the abrupt changes very quickly. Estimates were calculated for the given TV coefficients in (3.15). The standard deviations of the parameter estimates (with respect to the true parameters) and the statistics of the obtained results are presented in Table 3.2. The mean absolute error (MAE) of the parameter estimates, with respect to the corresponding true values, are also estimated and shown in Table 3.2. Compared with the normalized LMS estimates

and RLS estimates, the variance for the multi-wavelet basis functions block LMS method estimates is much smaller. The mean absolute error is defined by

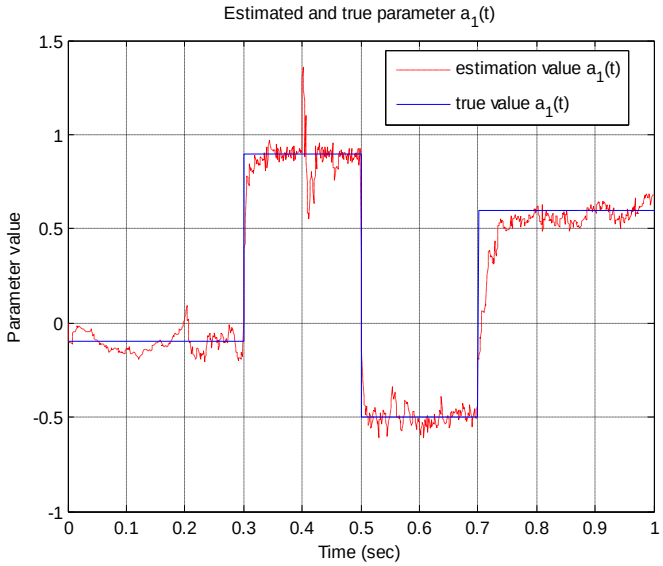
$$MAE = \frac{1}{N}\sum_{k=1}^{N}\left|\hat{a}(k) - a(k)\right|, \tag{3.16}$$

where $\hat{a}(k)$ represents the estimates of $a(k)$ in model (3.1), and N is the length of the data.

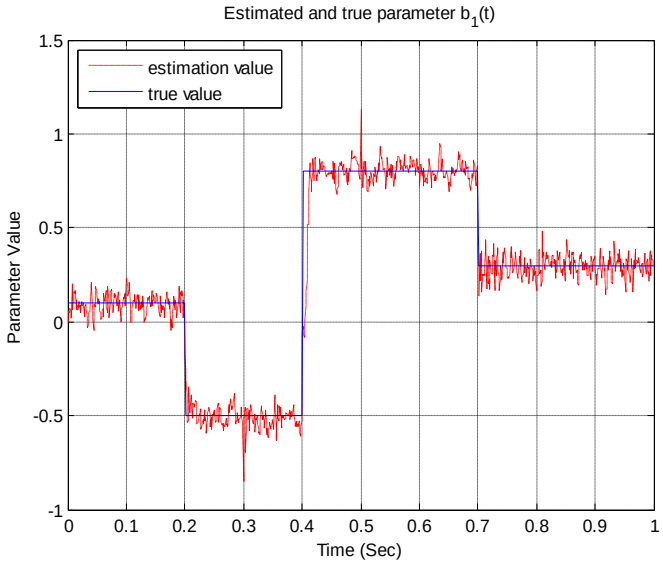
Table 3.2 A comparison of the model performance for Example 1 (SNR = 19.4024dB).

Approach	Estimated coefficient	MAE	Std
NLMS $(L=1,\ \mu=0.6)$	$\hat{a}_1(t)$	0.0616	0.1072
	$\hat{b}_1(t)$	0.0559	0.0968
RLS $(\mu=0.9)$	$\hat{a}_1(t)$	0.0993	0.2180
	$\hat{b}_1(t)$	0.0418	0.1070
BLMS $(L=2,\ \mu=1)$ with multi-basis method	$\hat{a}_1(t)$	0.0443	0.0901
	$\hat{b}_1(t)$	0.0557	0.0884

Table 3.2 statistically confirms the better performance of the block LMS multi-wavelet basis functions method. Compared with the traditional normalized LMS and RLS approaches, the simulation results above really show that the new method based on multi-wavelet basis functions proposed in this Chapter is more adaptive and possesses much better tracking ability in that it still can track the time-varying trend of the parameters even with noise contamination.



(a)

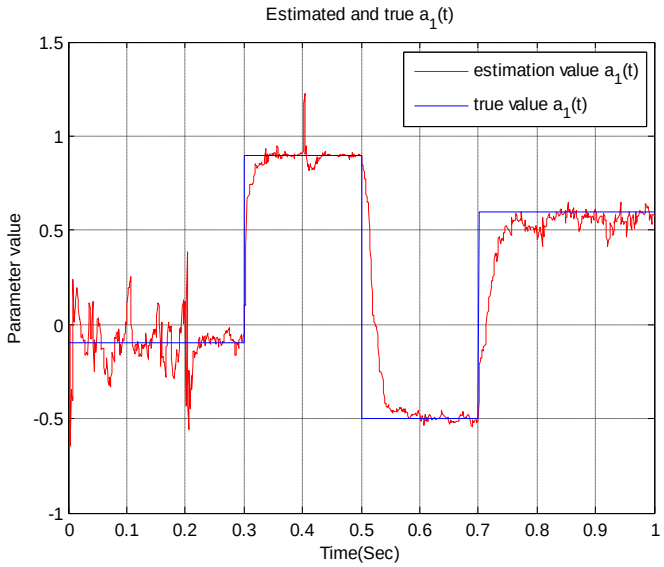


(b)

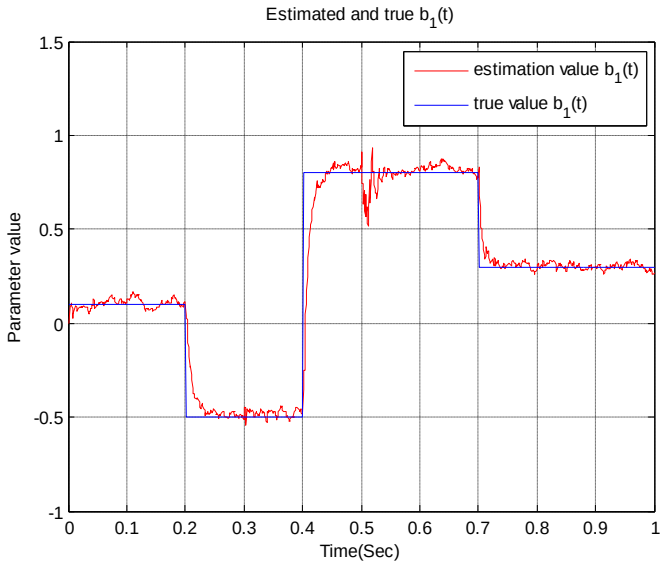
Figure 3.1 Time-varying parameter estimation using a normalized LMS approach with $L=1$, $\mu=0.6$ and SNR of 19.4024 dB for Example 1.

(a) Estimated and true parameter a_1 (t), and

(b) Estimated and true parameter b_1 (t).



(a)

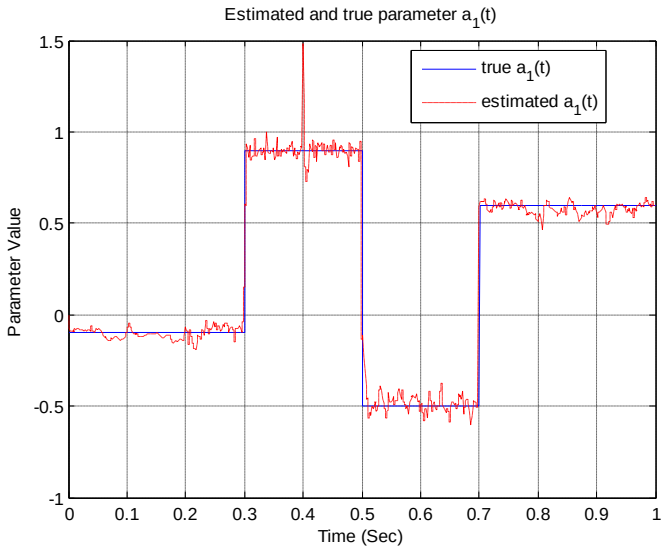


(b)

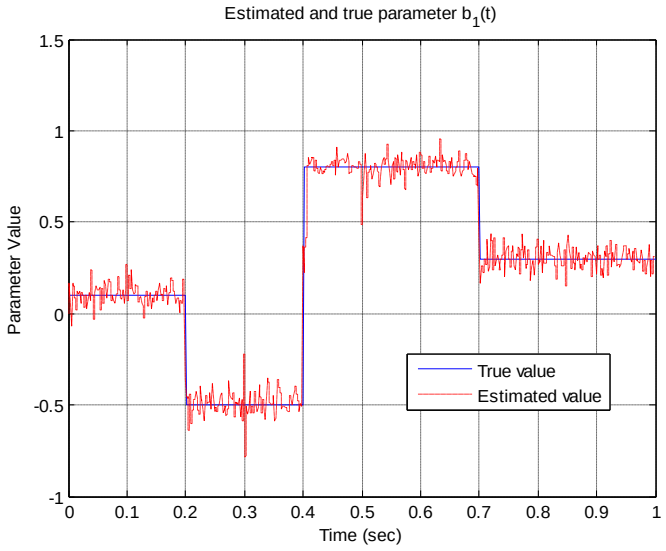
Figure 3.2 Time-varying parameter estimation using a RLS with forgetting factor $\mu = 0.9$ and SNR of 19.4024 dB for simulation example.

(a) Estimated and true parameter a_1 (t), and

(b) Estimated and true parameter b_1 (t).



(a)



(b)

Figure 3.3 Time-varying parameter estimation using the block LMS approach based on multi-wavelet basis functions with $L=2,\quad \mu=1$ and SNR of 19.4024 dB for simulation Example.

(a) Estimated and true parameter a_1 (t), and

(b) Estimated and true parameter b_1 (t).

3.6 Application—tracking of a mechanical system

The proposed modelling scheme has been applied to the analysis of a mechanical system to illustrate the application and tracking ability of the proposed multi-wavelet basis function method based TVARX modelling approach. The second-order continuous linear system equation subject to a general excitation $f(t)$ as formulated in (Ghanem & Romeo 2000) is given by

$$m(t)\ddot{y}(t)+c(t)\dot{y}(t)+k(t)y(t)=f(t), \tag{3.17}$$

The parameters $m(t)$, $c(t)$ and $k(t)$ represent the TV mass, damping and stiffness of the system, respectively. The properties and parameters of system are described in (Ghanem & Romeo 2000). An idealized model of damage characterized by an abrupt change in the stiffness is considered, which is to assess the suitability of the identification procedure to abrupt variations in the time-dependent parameters. The damped Mathieu equation is given by

$$m\ddot{y}(t)+c\dot{y}(t)+k(t)y(t)=u(t), \tag{3.18}$$

where $m(t)=1.0$, $c(t)=1.0$ are the parameter measuring the strength of the parametric excitation, and the excitation is given by $u(t)=A\sin(\omega_1 t)+A\sin(\omega_2 t)$, where $A=1.0$, $\omega_1=6.91\,rad/s$ and $\omega_2=5.65\,rad/s$, the step time used is $\Delta t=0.06s$, the initial conditions for the free vibrations are $x(0)=0.1$, $\dot{x}(0)=0$. The abrupt stiffness changes are given as $k(t)=k_0$ for $t<T/4$, $k(t)=0.75k_0$ for $T/4\le t<3/8T$, $k(t)=0.9k_0$ for $3/8T\le t<4.5/8T$, $k(t)=0.75k_0$ for $4.5/8T\le t<5/8T$, and $k(t)=0.9k_0$ for $5/8T\le t<T$. The value of parameter is $k_0=39.0$ and T is periodic. The difference equation for the system (3.18) using the approximations $\dot{y}(t)\approx$

$\frac{\left[y(t+\Delta t)-y(t)\right]}{\Delta t}$ and $\ddot{y}(t)\approx\frac{\left[\dot{y}(t+\Delta t)-\dot{y}(t)\right]}{\Delta t}$ can be approximated by

$$y(k\Delta t)=(2-\Delta t)\,y\big((k-1)\Delta t\big)+\big(-1+\Delta t-\Delta t^2k(t)\big)$$
$$\times y\big((k-2)\Delta t\big)+\Delta t^2u\big((k-1)\Delta t\big), \tag{3.19}$$

Eq. (3.19) can be considered as a discrete model approximation of (3.18). The selection of the order of B-splines, with the resolution level (scale index), is the same as in the previous example. Comparison with Eq. (3.1), the linear TVARX model (3.19) with $p=2$, and $q=1$ can be estimated by the proposed approach, with $\hat{a}_2(t)=-1+\Delta t-\Delta t^2k(t)$, then the stiffness estimation value of the time-varying parameter in model (3.19) $\hat{k}(t)$ can be represented by

$$\hat{k}(t)=\frac{\big(-1+\Delta t-\hat{a}_2(t)\big)}{\Delta t^2}, \tag{3.20}$$

The stiffness estimate of the TV parameter $k(t)$ with the new approach is given in Figure 3.4. Compared with the result given in (Ghanem & Romeo 2000), the parameter estimate here follows the true parameter variation very well and traces the abrupt changes quickly.

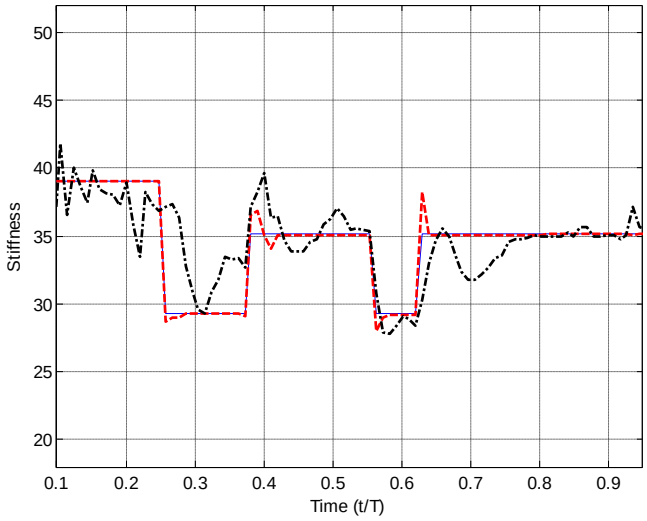


Figure 3.4 Time-varying abrupt stiffness parameter estimation $k(t)$ using the block LMS approach based on multi-wavelet basis functions with $L=2,\ \mu=0.78$. Solid (blue) line, the dashed (red) line and the dashed (black) line indicate the true value, estimate value (no noise) and estimate value with SNR (5.9735 dB), respectively.

3.7 Conclusions

Time-varying parameters in ARX models have been estimated using a new multi-wavelet basis function approach with a block LMS algorithm introduced in this Chapter, where the associated time-dependent coefficients are expanded using multi-wavelet basis functions. Parameter variations including abrupt or sharp changes have been considered. Performance measures of the estimated parameters have been calculated. The experimental results indicate that the new approach based on multi-wavelet basis functions with the block LMS algorithm gives much better results for fast and abrupt changing parameters than the method which uses the traditional normalized LMS algorithm or the RLS algorithm directly. Furthermore, from the results above, it can be concluded that time-varying

systems can be modelled using a TVARX and the identification problem of modelling fast and abrupt changing time-varying parameters is possible with good accuracy. The identification procedure has been shown to be effective in tracking time evolutions of the unknown parameters.

The wavelet method is especially powerful for nonstationary signal analysis. We advocate the use of a multi-wavelet basis functions to identify and model time-varying signals including mechanical systems because of the flexibility in capturing the signal characteristics at different scales.

The time-varying model with multi-wavelet basis functions is able to track and capture the parameter changes even those with jumps. In this time-varying modelling approach, the time varying coefficients were expressed using a combination of a number of wavelet basis functions. However, it should be pointed out that in many applications, not all these candidate wavelet basis functions need to be simultaneously involved in a same time varying coefficient approximation, some wavelet basis functions which play a more important role need be included in the expression, some other wavelet basis functions, however, may only play some little role and can be exclude from the expression. In the next Chapter, another novel multi-wavelet basis function expansions will be introduced and the orthogonal least squares (OLS) algorithms will then be applied to determine which wavelet basis functions should be included in the final approximation expression and which candidate wavelet basis functions should be eliminated from the dictionary.

Chapter 4

Time-Varying Model Identification for Time- Frequency Feature Extraction from EEG data

4.1 Introduction

Time-varying processes encountered in different engineering applications such as biomedical signal processing can be characterised by parametric representations (Chen & Chowdhury 2007; Chon et al. 2005; Wei & Billings 2002; Wei et al. 2010). Thus, the need to identify TV systems has naturally led to a growing interest in these areas. Parameter identification and modelling is now established based on the ARX model. The ARX model, which can match the structure of many real-world processes, is one of the most widely applied linear dynamic models. The popularity and wide application of the ARX model comes mainly from its easy-to-compute parameters (Burke et al. 2005; Nells 2001; Wei et al. 2009).

In Chapter 2 a solution to the identification of time varying system has been investigated by introducing a multi-wavelet basis function scheme. This Chapter will consider the time-varying modelling problem from another perspective, where a multi-wavelet basis function expansion approach is applied to EEG signal modelling and the orthogonal least squares (OLS) algorithm is then used to select the significant candidate model terms (formed by wavelet expansion) to form a parsimonious model structure and estimate the relevant model parameters.

It should be noted that the choice of basis functions can significantly affect the

performance of the parameter estimates. However, there is no guideline on how to choose the appropriate basis functions for a specific modelling problem. Conventionally, the basis functions have been chosen to be polynomials (including Chebyshev and Legendre types), prolate spheroidal sequences which are the best approximation to bandlimited functions (Chon et al. 2005; Niedzwiecki 2000; Wei & Billings 2002; Zou & Chon 2004) and wavelets that have a distinctive property of multi-resolution in both the time and frequency domains (Ansari-Asl et al. 2005; Cakrak & Loughlin 2001; Wei et al. 2010). In fact, each family of basis functions has its own properties of accuracy and tractability, for example, polynomial and Fourier basis functions can work well for most smoothly and slowly varying coefficients; wavelet basis functions, however, perform well for time-varying coefficients of abrupt changes or piecewise variations. Basis function expansion methods have been widely applied to solve various engineering problems. For example, a TVAR model can be expanded over a Fourier–Bessel (FB) series to constitute a feature vector for segmentation of the EEG signal, and then to find a simple model for the parametric representation of EEG signals (Pachori & Sircar 2008). A good choice of the basis functions should allow abruptly or rapidly changing parameters to be tracked.

Wavelets have distinctive approximation properties and are well suited for approximating general nonstationary signals (Chui 1992; Mallat 1989; Wei & Billings 2002; Wei & Billings 2006a; Wei et al. 2004b; Wei et al. 2010), and thus have been successfully applied to many areas including nonlinear signal processing and parametric identification (Adeli et al. 2003; Forte et al. 2008; Ng & Raveendran 2009; Tsatsanis & Giannakis 1993; Wei et al. 2004b). However, to our knowledge, not much work has been done to exploit the inherent approximation properties of wavelets to identify TV coefficient parameter estimation. The objective of this Chapter is to present a novel TVARX modelling approach, where the time-dependent coefficients are expanded using a finite set of multi-wavelet basis functions. Based on a multi-wavelet expansion scheme, a new method for

time-dependent parameter estimation is then proposed. The term ‘multi-wavelet’ here has a twofold meaning. Firstly, the TV coefficients of the ARX model are approximated using several types of wavelet basis functions (i.e. the TV parameter estimation involves multiple wavelets). Secondly, these wavelet basis functions are combined in a form of multi-resolution wavelet decomposition. The advantage of the proposed method, compared with a method involving only a single type of wavelets, is that the multi-wavelet expansion scheme is much more flexible in that it exploits the excellent properties of both non-smooth and smooth wavelet basis functions and thus can effectively track both rapid and slow variations of TV coefficients. In addition, the expansion of TV parameters onto multi-wavelet basis functions is more accurate and effective for dealing with nonstationary signal modelling than traditional power spectral estimation approaches and classical time-invariant parameter models.

Brain signals involved in neuro-physiological techniques such as EEG and MEG are inherently nonlinear and nonstationary processes which exhibit complex dynamics. Conventional time-domain analysis methods such as FFT and PSD estimation methods, where it is assumed that signals to be studied are stationary, are not ideal when applied to EEG data analysis. While traditional adaptive parameter estimation algorithms for example the RLS, LMS and Kalman filer approaches can be applied to track TV trends, these algorithms can often produce lagged tracking of TV parameters (for example in TVARX model estimation). The objective of this study is thus twofold: 1) to develop a TV modelling framework that exploits the excellent approximation property of wavelets so that the resultant time-varying models can track severely nonstationary processes with sharp changes in the system parameters; and 2) to apply the TV modelling algorithm to EEG data analysis where the proposed algorithm can provide almost instant tracking performance and the resultant time-frequency resolution produced by the identified model is extremely high; this is usually very important for feature extraction from EEG signals in both the time and frequency domains.

4.2 Problem formulation

4.2.1Time-Varying ARX Model and Multi-wavelet Coefficient Expansions

The TVARX(p,q) model for a single-input/output system can be represented as

$$y(t)=\sum_{i=1}^{p}a_i(t)\,y(t-i)+\sum_{l=1}^{q}b_l(t)\,u(t-l)+e(t), \tag{4.1}$$

where t is the time instant or sampling index of the signal $y(t)$, $y(t-i)$ and $u(t-l)$ are the measured response, respectively. $a_i(t)$ and $b_l(t)$ are the TV coefficient functions to be determined in the model; the term $e(t)$ is the residual error accommodating the effects of measurement noise, and modelling noise that can be viewed as a stationary white noise sequence with zero mean and variance σ_e^2. The proposed method is to expand the TV parameters $a_i(t)$ and $b_l(t)$ onto multi-wavelet families cardinal B-splines basis functions, $\left\{\phi_k^{(m)}:\ m=3,4,5;\ k\in\Gamma_m\right\}$, such that the following expression hold:

$$a_i(t)=\sum_{k\in\Gamma_\eta}\alpha_{i,k}^{(\eta)}\phi_k^{(\eta)}\left(\frac{t}{N}\right)+\sum_{k\in\Gamma_r}\alpha_{i,k}^{(r)}\phi_k^{(r)}\left(\frac{t}{N}\right)+\sum_{k\in\Gamma_s}\alpha_{i,k}^{(s)}\phi_k^{(s)}\left(\frac{t}{N}\right),$$

$$b_l(t)=\sum_{k\in\Gamma_\eta}\beta_{l,k}^{(\eta)}\phi_k^{(\eta)}\left(\frac{t}{N}\right)+\sum_{k\in\Gamma_r}\beta_{l,k}^{(r)}\phi_k^{(r)}\left(\frac{t}{N}\right)+\sum_{k\in\Gamma_s}\beta_{l,k}^{(s)}\phi_k^{(s)}\left(\frac{t}{N}\right), \tag{4.2}$$

where $\alpha_{i,k}$ and $\beta_{l,k}$ represent the expansion parameters, $\Gamma_m=\left\{k:\ -m\le k\le 2^j-1\right\}$ for $m=3,4,5$, $j=3$ is the wavelet scale, $\eta=3$, $r=4$, and $s=5$, $t=1,2,\cdots,N$, and N is the number of observations of the measurement data, respectively. Substituting (4.2) into (4.1), it yields,

$$y(t)=\sum_{i=1}^{p}\left\{\sum_{k\in\Gamma_m}\alpha_{i,k}^{(m)}\left(\phi_k^{(m)}\left(\frac{t}{N}\right)y(t-i)\right)\right\}+\sum_{l=1}^{q}\left\{\sum_{k\in\Gamma_m}\beta_{l,k}^{(m)}\left(\phi_k^{(m)}\left(\frac{t}{N}\right)u(t-l)\right)\right\}, \quad (4.3)$$

From (4.3), the original TVARX model in Eq. (4.1) has now been converted into a time invariant (TIV) regression model with respect to the time invariant coefficients $\alpha_{i,k}^{(m)}$ and $\beta_{l,k}^{(m)}$. In this Chapter, cardinal B-splines wavelets, which have been proved to have several excellent properties including orthogonality, support, regularity and time-frequency localization in a window which enables the operation of the multi-resolution decomposition to be much more convenient, are considered and will be employed for TV parameter expansion. Detailed discussion how to build the associated multi-wavelet model using B-splines can be found in (Wei & Billings 2002) and (Wei et al. 2010).

4.2.2 Time-Dependent Spectrum Estimation

Eq. (4.3) can be solved by using linear least squares algorithms. Let $\hat{a}_i(t)$, $\hat{b}_l(t)$ be the estimates of $a_i(t)$ and $b_l(t)$, and $\hat{\sigma}_e^2$ is the estimate of σ_e^2. The time-dependent spectral function associated to the TVARX model in Eq. (4.1) is defined as,

$$H(f,t)=\left|\frac{\sum_{n=1}^{q}\hat{b}_l(t)e^{-j2\pi lf/f_s}}{1-\sum_{i=1}^{p}\hat{a}_i(t)e^{-j2\pi if/f_s}}\right|^2, \quad (4.4)$$

where $j=\sqrt{-1}$, and f_s is the sampling frequency. Note that the spectral function (4.4) is continuous with respect to the frequency f and thus can be used to produce spectral estimates at any desired frequency up to the Nyquist frequency $f_s/2$. The frequency resolution is not infinite, but is determined by the underlying model order and the associated parameters.

4. 3 Model Identification and Parameter Estimation

In general, the estimation of TIV system parameters is formulated as an overdetermined problem. Then the least squares solution is the optimal estimate of the parameters in the sense of minimum residual error. However, if the parameters are time-varying, the problem of parameter estimation becomes underdetermined, and it is much more difficult to find the 'best' solution. Expanding the TV parameters onto a linear or nonlinear combination of a set of basis functions can solve the underdetermined problem. Consequently, the parameter estimation of unknown variables can be reduced to a set of constant coefficients of the basis functions. However, the multi-wavelet expansion model (4.3) involves a large number of candidate model terms that may be highly correlated. The resultant parameter estimates may be over-fitted. Modelling experience suggests that most of the candidate model terms can be removed from the model, and that only a small number of significant model terms are needed to provide a satisfactory representation for most linear and nonlinear dynamical systems. Many approaches have been introduced to eliminate the possible linear dependency of candidate model terms by selecting best bases, for example, Kaipio and Karjalainen (1997) introduced a principal-component-analysis (PCA)-type approximation scheme to select the 'optimal basis'. The mutual correlation of the coefficients is also taken into account in their approach.

In this work, TV coefficients are expanded by multi-resolution cardinal B-splines wavelet series, and then the forward OLS algorithm (Billings et al. 1989; Billings & Wei 2007; Chen & Billings 1989; Wei & Billings 2008), which have been proven to be a very effective to deal with multiple dynamical regressions problems, is applied to determine the forms in model (4.3). The orthogonal least squares (OLS) algorithm is one of the most efficient techniques that can be used to detect the

model structure. The OLS algorithm orthognalizes all the regression terms in Eq. (4.3), step by step, in a forward stepwise manner, by introducing an auxiliary orthogonal model. The error reduction ratio (ERR) values can then be applied as a measure of the significance of each candidate model terms. The OLS algorithm has been widely applied to select the model structure selection for nonlinear system identification (Billings et al. 1989; Chen et al. 1989) and has already become a standard algorithm for system identification, nonlinear function approximation and neural network training (Harris et al. 2002; Haykin 2003). Detailed discussions and derivation of the procedure of the forward OLS can be found in (Billings et al. 1989; Billings & Wei 2007; Chen & Billings 1989; Wei & Billings 2008; Wei et al. 2004b). The TV coefficients $a_i(t)$ and $b_n(t)$ in Eq. (4-1) can be recovered by the resultant estimates from model (4.3). A brief introduction of the OLS algorithm is given in the Appendix 4.1

4.4 Simulation example

Consider a TVARX (2, 2) model below

$$y(t)=a_1(t)y(t-1)+a_2(t)y(t-2)+b_1(t)u(t-1)+b_2(t)u(t-2)+e(t), \quad (4.5)$$

where $e(t)$ is zero-mean Gaussian white noise. The TV parameters in Eq. (4.5) are given by:

$$a_1(t)=\begin{cases} 0.32\cos\left(1.5-\cos\left(\frac{4\pi t}{N}+\pi\right)\right), & 1\le t\le \frac{N}{4} \\ 0.32\cos\left(3-\cos\left(\frac{4\pi t}{N}+\frac{\pi}{2}\right)\right), & \frac{N}{4}+1\le t\le \frac{3N}{4} \\ 0.32\cos\left(1.5-\cos\left(\frac{4\pi t}{N}+\pi\right)\right), & \frac{3N}{4}+1\le t\le N \end{cases},$$

$$a_2(t)=0.4\cos\left(\frac{4\pi t}{N}\right), 1\le t\le N,$$

$$b_1(t)=\begin{cases}0.65, & 1\le t\le \dfrac{N}{4}\\ -0.5, & \dfrac{N}{4}+1\le t\le \dfrac{N}{2}\\ 0.65, & \dfrac{N}{2}+1\le t\le \dfrac{3N}{4}\\ -0.5, & \dfrac{3N}{4}+1\le t\le N\end{cases}, \tag{4.6}$$

$$b_2(t)=0.6,\ 1\le t\le N,$$

where the length of data N is 512. Model (4.5) was simulated by setting the input $u(t)$ as a Pesudo-Random Binary Sequence (PRBS) (Leontaritis & Billings 1987). The variance of the noise $e(t)$ was chosen to be 0.04, and this made the signal-to-noise ratio (SNR) to be around 13 dB. Both the input and the associated output sequences were recorded and were used for subsequent model estimation.

Figure 4.1 compares three different methods, that is, the RLS algorithm, the RLS algorithm with B-splines basis functions, and the OLS algorithm with B-splines basis functions. Panel (a) shows the results using the RLS estimation algorithm (forgetting factor (ff) 0.92) (Ljung & Gunnarsson 1990). Panel (b) gives the results of the RLS (ff: 0.9998, using B-splines wavelets and selecting scale index $j=3$) algorithm and Panel (c) shows the OLS identification results (using B-spline wavelets and selecting scale index $j=3$). Obviously, The RLS approach attains smooth but relatively poor estimates that cannot track the rapidly changing TV parameters, the parameter estimates are underdetermined. The RLS approach with B-splines obtains irregular estimates with large variances (over-fitted), however, compared with RLS method, the resultant estimates from the RLS method with B-splines can track the sharp changes of the TV parameters. These interesting results have been verified by Li et al. (2011a). The OLS method with B-splines appears to outperform the RLS approach and the RLS approach with B-splines. The results using the OLS approach with B-splines is impressive because it is able to track three quite different waveforms: the constant value, an abrupt change, and the sinusoidal waveform. The proposed method (the OLS with B-splines) can attain

smooth estimates while providing rapid tracking. The mean absolute error (MAE), normalized root mean squared error (RMSE) and the standard deviations (*Std*) of the parameter estimates (with respect to the true parameters) are estimated and shown in Table 4.1.

Compared with the RLS approach and the RLS approach with B-splines estimates, Table 4.1 statistically confirms that the MAE, RMSE and Std estimates produced by the OLS approach with B-splines yield smallest. The MAE and RMSE are both defined by

$$MAE = \frac{1}{N}\sum_{t=1}^{N}\left|\hat{a}(t) - a(t)\right|, \tag{4.7}$$

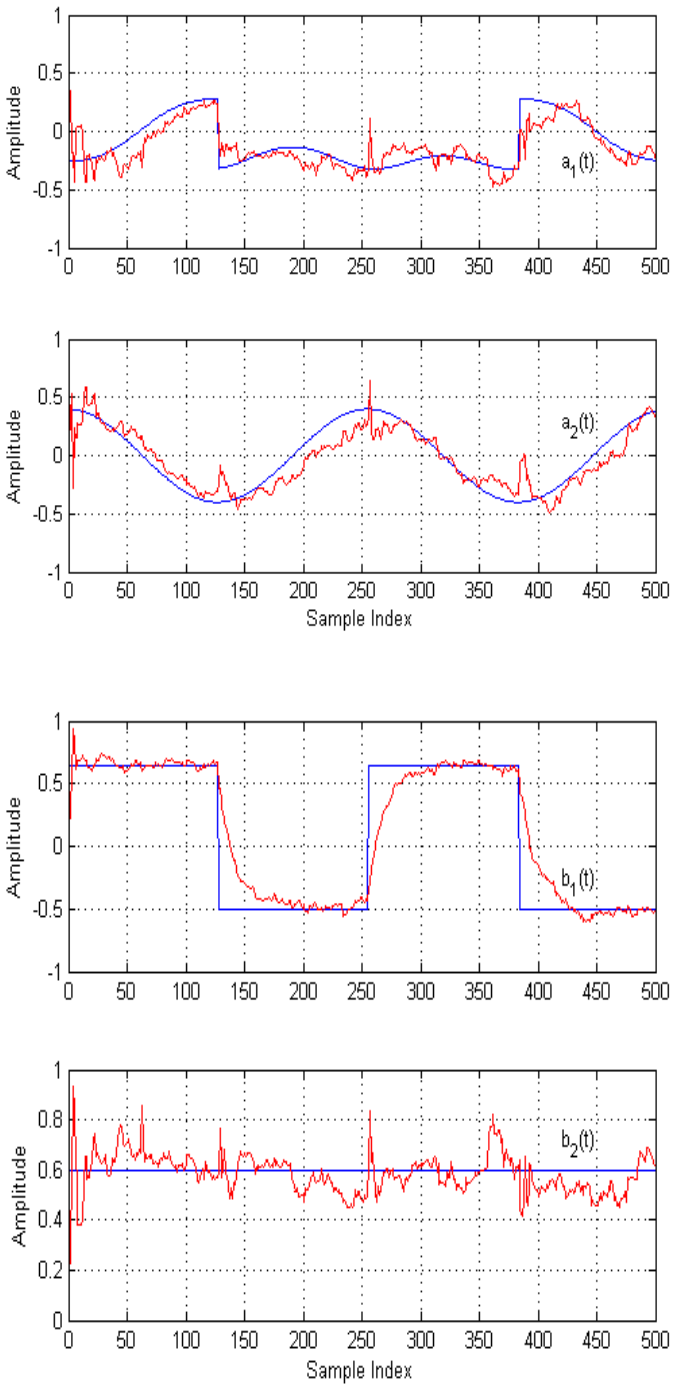
$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\frac{\left\|\hat{a}(t) - a(t)\right\|^2}{\left\|a(t)\right\|^2}}, \tag{4.8}$$

where $\hat{a}(t)$ represents estimates of coefficients $a(t)$ in the TVARX model (4.3), and N is the length of the data set.

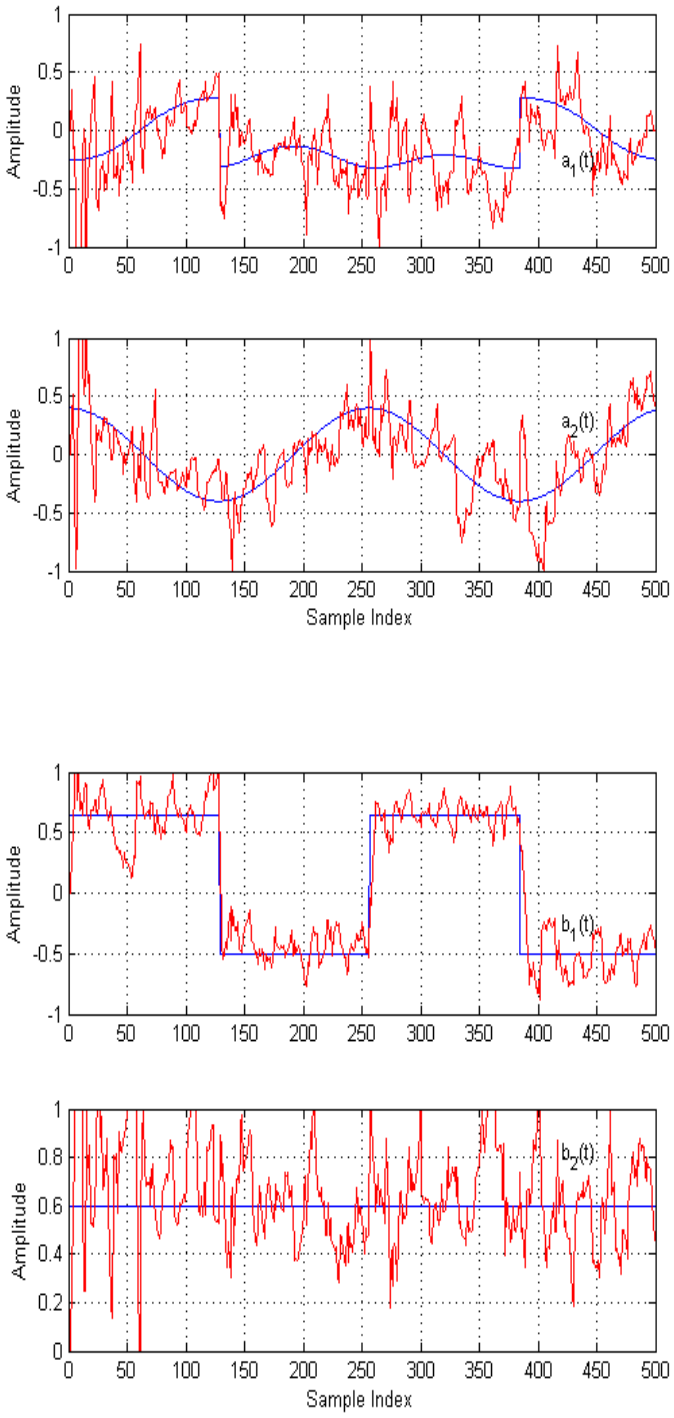
Table 4.1 A comparison of the model performance for TVARX (2, 2) model with SNR 13 dB.

Approach	Estimated coefficients	MAE	RMSE	Std
RLS $(\mu = 0.92)$	$\hat{a}_1(t)$	0.0917	2.3104	0.1199
	$\hat{a}_2(t)$	0.1030	1.8667	0.1292
	$\hat{b}_1(t)$	0.1080	1.2315	0.2172
	$\hat{b}_2(t)$	0.0627	0.7260	0.0863
RLS with B-splines $(\mu = 0.9998)$	$\hat{a}_1(t)$	0.2045	3.0639	0.2623
	$\hat{a}_2(t)$	0.2047	1.9951	0.2746
	$\hat{b}_1(t)$	0.1411	1.3973	0.2084
	$\hat{b}_2(t)$	0.1803	1.0434	0.2900
OLS with B-splines	$\hat{a}_1(t)$	0.0893	2.1750	0.1112
	$\hat{a}_2(t)$	0.0614	1.3520	0.0865
	$\hat{b}_1(t)$	0.0642	0.8638	0.1133
	$\hat{b}_2(t)$	0.0246	0.2250	0.0305

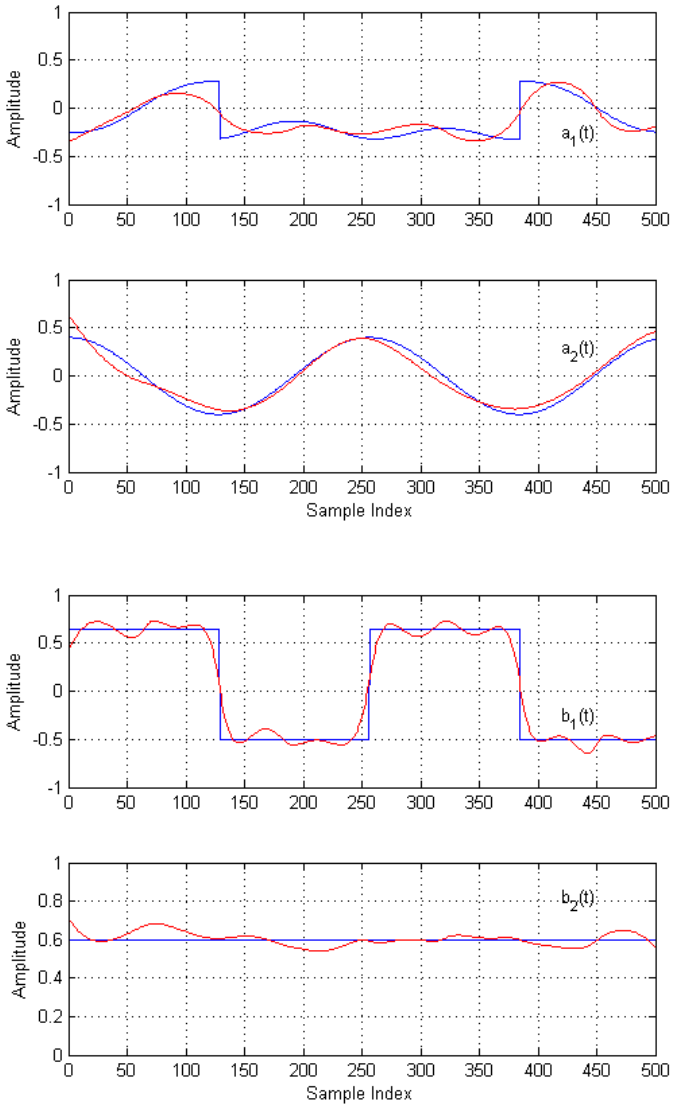
where μ represents the forgetting factor.



(a) RLS method



(b) RLS method with B-splines



(c) OLS method with B-splines

Figure 4.1 One implementation of the TVARX (2, 2) system identification results with a SNR of 13 dB using the different approaches. Blue curve represents the true value of the TV parameters; red curve indicates the estimation value of the TV parameters.

(a) the RLS method;

(b) the RLS method with B-splines;

(c) the OLS method with B-splines.

4.5 Application—EEG data modelling and analysis

4.5.1 Time-varying ARX model for EEG

The proposed TVARX modelling scheme has been applied to analyse dynamic relationships from EEG recordings to illustrate the application of the proposed multi-wavelet basis function method based on TV parametric modelling. Scalp EEG signals are synchronous discharges from cerebral neurons detected by electrodes attached to the scalp. The EEG signals considered here were recorded, with the same 32-channel amplifier system. An XLTEK 32 channel headbox (Excel-Tech Ltd) with the international 10-20 electrode placement system was used in the Sheffield Teaching Hospitals NHS Foundation Trust, Royal Hallamshire Hospital, UK. The sampling frequency of the device was 500 Hz. Andrzejak et al. (2001) has discussed in detail dynamical properties of brain electrical activity from different extracranial and intracranial recording regions and from different physiological and pathological brain states. The central objective of EEG signal processing here is to propose an empirical and data-based modelling framework that can produce an accurate but simple description of the dynamical relationships between different recording regions during brain activity. This is a complicated black box system where the true model structure is unknown, and thus, needs to be identified from available experimental data. Simulation examples have shown that proposed approach is more capable of tracking severely nonstationary processes with sharp changes, compared to the traditional adaptive method including RLS algorithm. The proposed time varying modelling algorithm can then be applied to EEG data analysis to extract the time-frequency feature and help clinicians to interpret EEG signals. As an example, the symmetrical two channels (F3, located over the left superior frontal area of the brain, and F4 located over the same area on the right) of EEG recorded from a patient with an absence of seizure epileptic

discharge was investigated. Channel F3 was treated as the input, denoted by $u(t)$, and Channel F4 was treated as the output, denoted by $y(t)$, note that Channel F3 is the signal input and Channel F4 is the signal output, the main reason is that the phase of Channel F4 is related to the phase of Channel F3. The objective is to investigate, from the available Channel F3 and Channel F4 recordings, if an identified TVARX model is suitable to describe the dynamical characteristics by using the time-dependent spectrum analysis approach. A data set, consisting of 3500 input-output data point pairs of EEG signals representing one seizure epoch recorded for a total of 7 seconds with a sampling rate of 500 Hz, was analysed in this example. The plots of the associated samples are shown in Figure 4.2.

Similar to the simulation example given in section 4.3, the third, fourth and fifth order B-splines were adopted to establish TVARX models for the EEG recordings. Several TVARX models with different model orders were estimated using the OLS approach with B-splines, the classical generalized cross-validation (GCV) criteria (Billings et al. 2007) suggested that the model order can be chosen to be $p=4$ and $q=3$ when using the B-splines as building blocks to represent the time-varying coefficients in the TVARX model.

The time-varying coefficients estimated $a_i(t)$ with $i=1,2,\cdots,4$ and $b_l(t)$ with $l=1,2,3$ are depicted in Figure 4.3. Figure 4.4 shows the recovered signal, recovered by the TVARX model from the estimated time-varying coefficients $a_i(t)$ and $b_l(t)$. The topographical diagram of the time-dependent spectrum estimated from the TVARX (4, 3) model is shown in Figure 4.5, and the 2-D image diagram and the contour plot of the time-dependent spectrum produced from the 3-D topographical diagram are given in Figure 4.6.

From Figure 4.5 and Figure 4.6, the distribution scale of the power spectrum of the EEG signal considered here is mainly from zero to around 18 Hz. Two frequency bands can clearly be observed as: 1) the low frequency band (about the 3 Hz, namely, a spike at 3-Hz); 2) around 18 Hz represents the high frequency band component. The contour plot of the time-dependent spectrum given in Figure 4.6(b) clearly reflects the distribution of these frequency components along with the time course. It is clear that the variations of the time course signals can be observed from the contour diagram of the transient spectrum. For instance, the power spectrum is mainly distributed by a 3-Hz spike frequency component during the period from 5 to 6s, while the high frequency (around 18 Hz) activity is dominated by the time course from 0.2 to 0.3s. Any time-invariant parametric modelling framework such as the commonly applied ARX models cannot attain these properties which are only possessed by the TVARX model proposed.

Note that in order to avoid any distortion of the EEG signals, no filtering was applied to pre-process the raw EEG signals here (and in the example below) apart from the notch frequency (50Hz) being removed.

4.5.2 Time-varying AR model for EEG

This example presents some illustrations to demonstrate the applicability of the proposed modelling framework for characterising epileptic seizure EEG signals by using time-varying AR models. Again, channel C3 of the EEG signal of a subject for an epileptic seizure activity lasting 20 seconds, was considered. A total of 10000 samples, recorded with a sampling rate of 500Hz, are shown in Figure 4.7.

Similar to the previous example, the third, fourth and fifth order B-splines were adopted to establish TVAR models for the EEG recordings. Numerical experiments have suggested that a TVAR model of order $p = 4$ is appropriate for representing

the EEG recordings here. The topographical diagram of the time-dependent spectrum estimated from the TVAR (4) model is shown in Figure 4.8, and the 2-D image diagram and the contour plot of the time-dependent spectrum produced from the 3-D topographical diagram are given in Figure 4.9.

From Figures 4.8 and 4.9, the power spectrum of the EEG signal with an epileptic seizure activity of 20 seconds is mainly distributed in the frequency range from zero to 10 Hz. Three frequency bands, i.e. Delta band (0-4 Hz), Theta band (4-7 Hz) and Alpha band (8-12 Hz), can clearly be observed as follows: 1) the low frequency band (around 3 Hz); 2) the frequency band that is centralized around 6 Hz; 3) around 10 Hz representing the high frequency band component. The image and contour plot of the time-dependent spectrum shown in Figure 4.9 clearly reflects the distribution of these frequency components along with the time course. It is clear that the variations of the time course signals can be observed from the contour diagram of the transient spectrum. For example, the power spectrum is mainly dominated by a 3-Hz frequency component during the period of the first 2 seconds, 7 to 8 seconds, and 11 to 12 seconds, respectively, while the high frequency (around 10 Hz) activity is dominated by the time course from 16 to 18 seconds. Any time-invariant parametric modelling framework such as the commonly applied conventional AR models cannot reveal these properties.

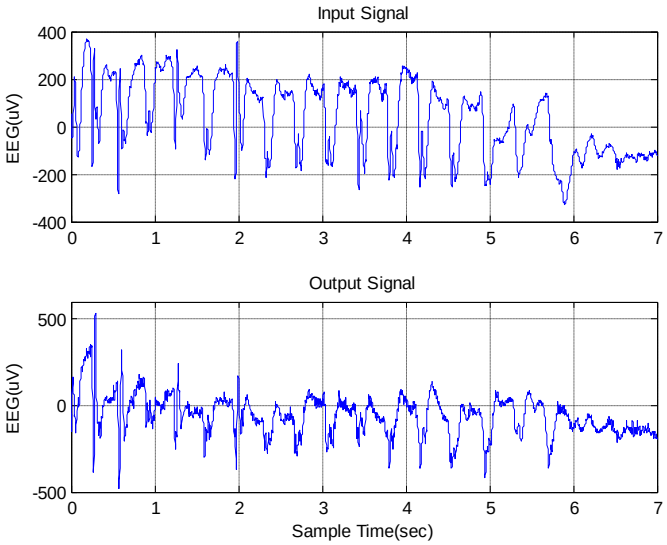


Figure 4.2 The EEG recordings (F3 Channel: Input signal, F4 Channel: Output signal), for a seizure activity of a patient, recorded over 7 seconds, with a sampling rate of 500 Hz.

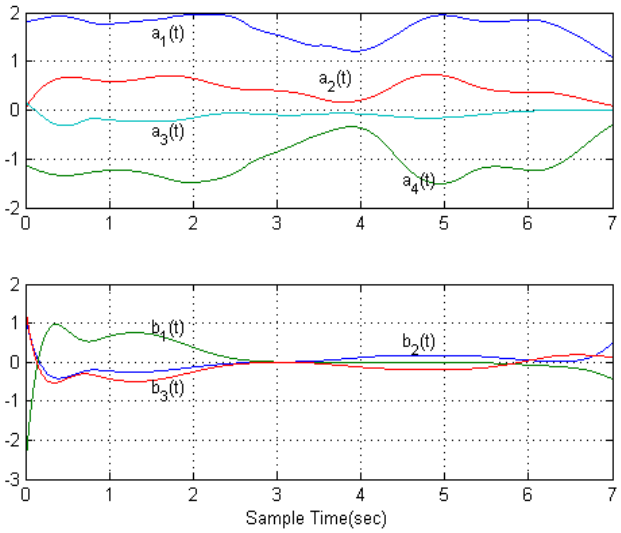


Figure 4.3 Estimates of the time-varying coefficients $a_i(t)$ for $i=1,2,3,4$ and $b_n(t)$ for $n=1,2,3$ for the EEG signal.

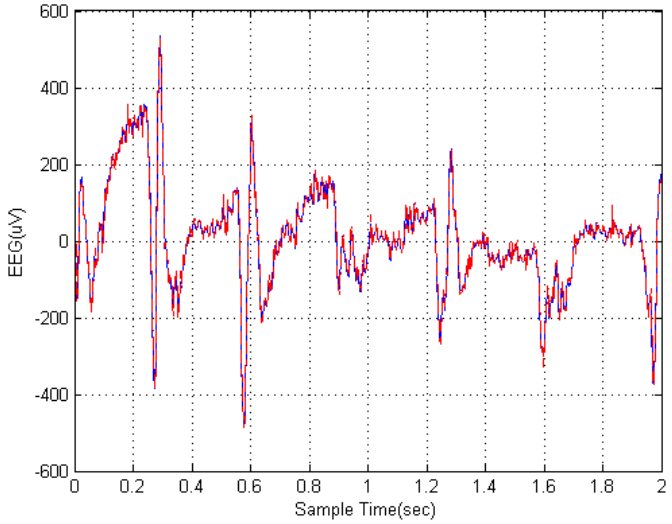


Figure 4.4 A comparison between the recovered signal from the identified TVARX (4, 3) model and the original observations for the EEG signal. Solid (blue) line represents the observations and the dashed (red) line represents the signal recovered from the TVARX (4, 3) model with One-Step-Ahead prediction. For a clear visualization only the data points of the period from 0 to 2 seconds are displayed.

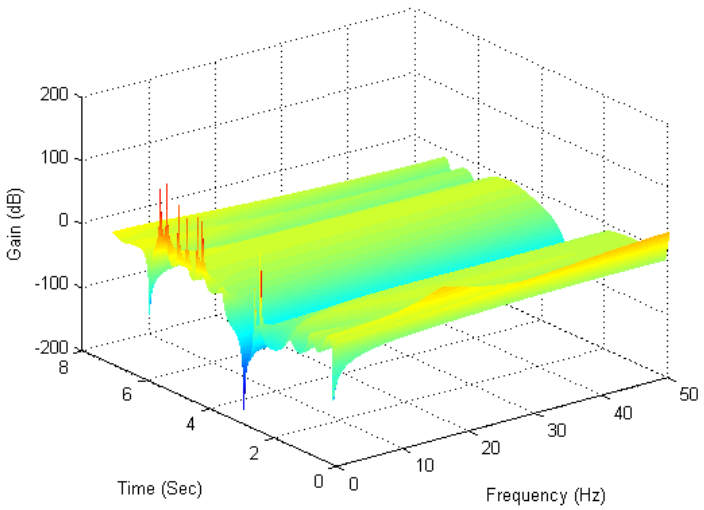
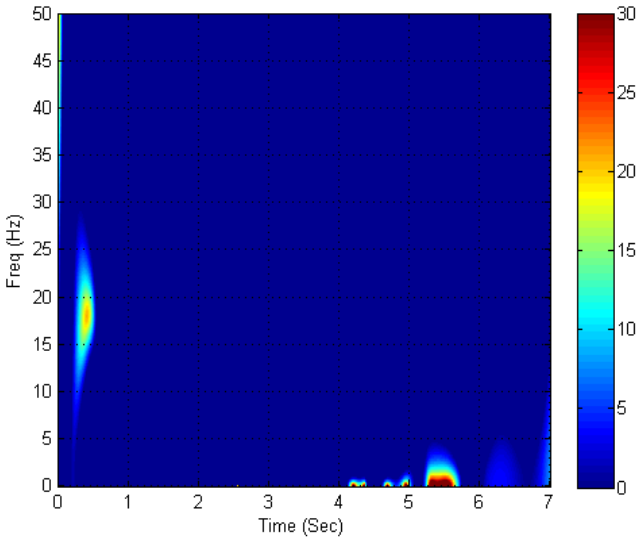
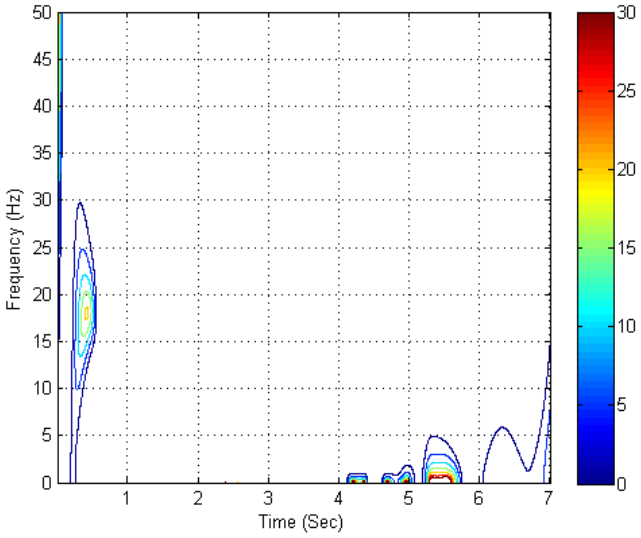


Figure 4.5 The 3-D topographical map of the time-dependent spectrum estimated from the TVARX model for the EEG signal.



(a)



(b)

Figure 4.6 The 2-D image and the contour diagram of the time-dependent spectrum produced by the 3-D topographical map for the EEG signal given in Figure 4.2.

(a) the 2-D image;

(b) the contour diagram.

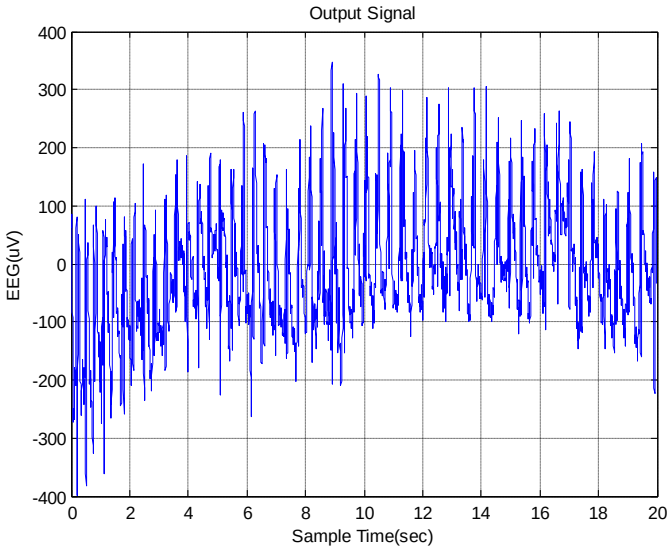


Figure 4.7 The EEG recordings of C3 Channel, for a seizure activity of an epileptic patient, recorded over 20 seconds, with a sampling rate of 500 Hz.

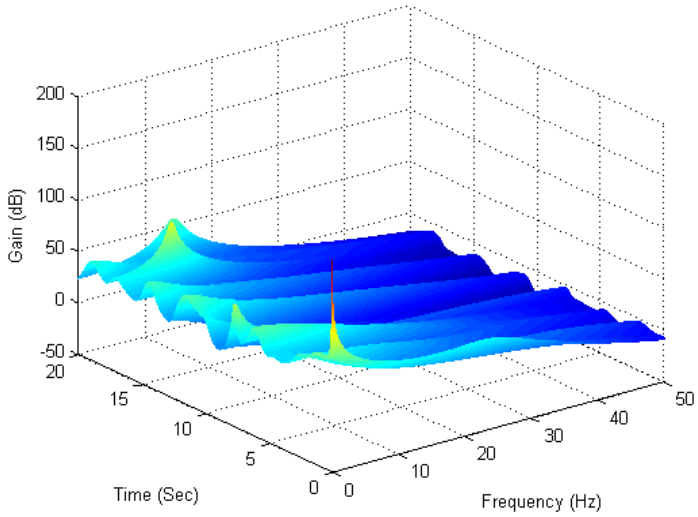
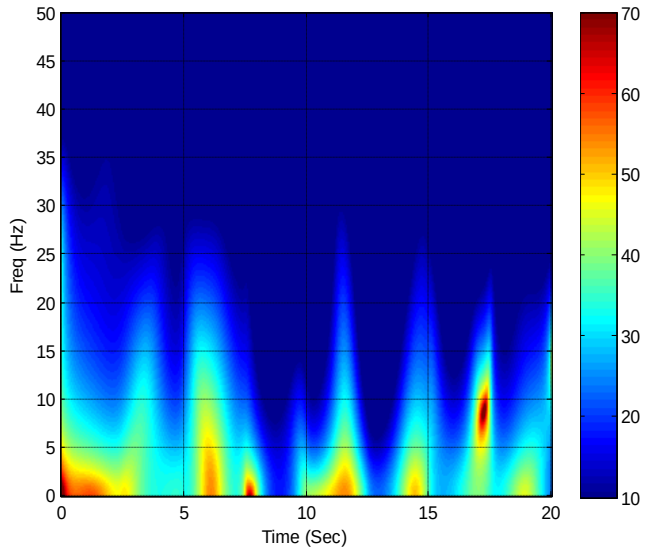
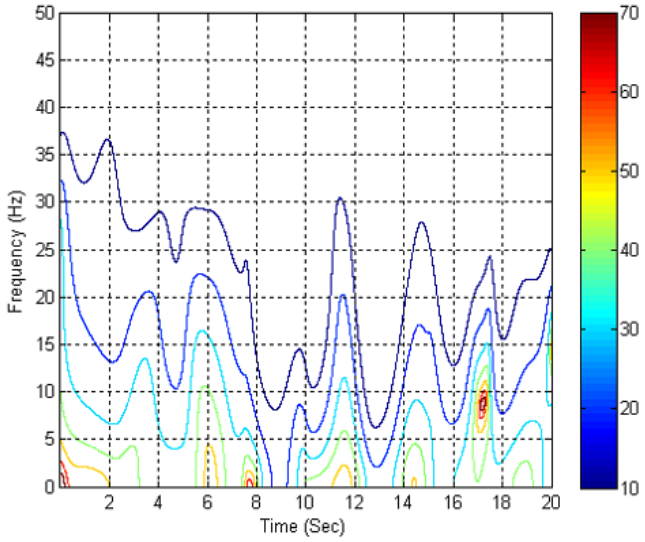


Figure 4.8 The 3-D topographical map of the time- dependent spectrum estimated from the TVAR (4) model for the EEG signal given in Figure 4.7.



(a)



(b)

Figure 4.9 The 2-D image and the contour diagram of the time-dependent spectrum produced by the 3-D topographical map for the EEG signal given in Figure 4.7.

(a) The 2-D image;

(b) The contour diagram.

4.6 Conclusions

In this Chapter, a novel TV parametric modelling approach has been presented based on time-dependent coefficients approximated multi-wavelet basis functions to account for the transient spectrum information, whereas the basis functions involved are locally defined. Using the method given in Chapter 3, a time varying multi-wavelet basis function expansion model can be constructed. The orthogonal least squares algorithm is then applied to select significant model terms (wavelet basis functions) from the initially pre-specified model set. Finally, the time-varying model parameters can be recovered by using the selected significant wavelet basis functions.

The time-dependent spectrum based on TVARX and TVAR model, with multi-wavelet basis functions, can reflect the global frequency behaviour of the signal and to reveal the local variations of the signal along the time course. One advantage of the proposed model, compared with traditional time-invariant models, is that it can capture much more transient information of the inherent nonstationary dynamics of the associated processes.

It is believed that the proposed time-varying modelling framework, coupled with results obtained from some other useful nonparametric methods for example higher-order statistics approach (Zhou et al. 2008), can promise significant new results which can reveal new and important features buried in EEG signals, which can in turn enable further studies and analysis, for example for disease detection and diagnosis. In addition, it is well known that EEG signals are nonlinear processes, and thus linear models may not be sufficient to represent such nonlinear processes; nonlinear representations should be more suitable for EEG data modelling and analysis. Time varying nonlinear modelling and identification of EEG signals will be discussed in the next Chapter.

Appendix 4.1 Forward Orthogonal Least Squares algorithm

In this subsection, a brief introduction of the OLS algorithm is given as follows. Consider the linear-in-the-parameters polynomial model

$$y(t)=\sum_{i=1}^{M}\theta_i\phi_i(t)+e(t), \tag{4.1.1}$$

where θ_i are unknown parameters, $\phi_i(t)$ are model terms. Assume the data length t is N. The matrix form of Eq. (4.1.1) can be represented by

$$Y=\Phi\Theta, \tag{4.1.2}$$

The matrix Φ is often referred to as the regression matrix. The regression matrix can be orthogonally decomposed as

$$\Phi=WA, \tag{4.1.3}$$

where A is an $M\times M$ unit upper triangular matrix and

$$W=\left[w_1,\cdots,w_M\right], \tag{4.1.4}$$

is a $N\times M$ matrix with orthogonal columns that satisfy

$$W^TW=D, \tag{4.1.5}$$

and D is a positive diagonal matrix $D=diag\left[d_1,d_2,\cdots,d_M\right]$ with $d_i=\langle w_i,w_i\rangle$, where the symbol $\langle .,.\rangle$ denotes the inner product of two vectors, that is

$$\langle w_i,w_i\rangle=\sum_{t=1}^{N}w_t(t)w_i(t), \tag{4.1.6}$$

Eq. (4.1.2) can now be expressed as

$$Y=\left(PA^{-1}\right)\left(A\Theta\right)=WG, \tag{4.1.7}$$

Rewrite Eq. (4.1.7) as

$$Y=\sum_{i=1}^{M}w_ig_i\ , \tag{4.1.8}$$

and calculate the inner product $\langle w_i,Y\rangle$, substituting Y by Eq. (4.1.8) yields

$$\langle w_i, Y \rangle = \left\langle w_i, \sum_{i=1}^{M} w_i g_i \right\rangle = \sum_{i=1}^{M} \langle w_i, w_i g_i \rangle = g_i \langle w_i, w_i \rangle, \tag{4.1.9}$$

Calculate the inner product $\langle Y, Y \rangle$ from the Eq. (4.1.8)

$$\langle Y, Y \rangle = \left\langle \sum_{i=1}^{M} w_i g_i, Y \right\rangle = \sum_{i=1}^{M} \langle w_i g_i, Y \rangle = \sum_{i=1}^{M} g_i^2 \langle w_i, w_i \rangle, \tag{4.1.10}$$

Dividing both sides of Eq. (4.1.10) by $\langle Y, Y \rangle$, then yields

$$1 = \sum_{i=1}^{M} \frac{g_i^2 \langle w_i, w_i \rangle}{\langle Y, Y \rangle}, \tag{4.1.11}$$

The error reduction ratio ERR_i due to w_i is can be presented by

$$ERR_i = \frac{g_i^2 \langle w_i, w_i \rangle}{\langle Y, Y \rangle}, \tag{4.1.12}$$

From the Eq. (4.1.9),

$$g_i = \frac{\langle w_i, Y \rangle}{\langle w_i, w_i \rangle}, \tag{4.1.13}$$

The error reduction ratio often provides a simple and effective means for select a subset of significant terms from a large number of candidate terms in a forward regression manner. A term can be selected if it produces the largest value of ERR_i among the rest of the candidate terms. The selection procedure will be terminated when

$$1 - \sum_{i=1}^{M_0} ERR_i < \varepsilon, \tag{4.1.14}$$

where ε is a desired tolerance, and this leads only to a subset model of M_0 terms $(M_0 << M)$. The detail procedure can be seen Billings and Chen (1989).

Chapter 5

Identification of nonlinear time-varying systems using an online sliding-window and common model structure selection (CMSS) approach with applications to EEG

5.1 Introduction

In Chapters 3 and 4, time varying linear model identification methods were investigated. However, many processes in engineering systems and biomedical neuroscience exhibit both time-varying and nonlinear behaviours. The identification of mathematical models of dynamical nonlinear systems is vital in many fields. During recent years, much attention has been devoted to the problem of identification of time-varying systems. In many practical cases, the system parameters are unknown and are time varying. When the system is given in state-space form, a classical approach consists of applying Kalman filter based algorithms for estimation of time-varying parameters (Morbidi et al. 2008). The application of the sliding-window recursive least squares (SWRLS) algorithm to the estimation of nonlinear system parameters often requires the nonlinear model outputs to be expressed linearly in terms of the unknown parameters. A discussion about performance of SWRLS identification and related adaptive control schemes can be found in (Choi & Bien 1989). Schilling et al. (2001) introduced parameter estimation methods based on a radial basis functions (RBF) neuronal predictor.

Although different approaches have been investigated for nonlinear system state

estimation in (Chen & Billings 1989; Choi & Bien 1989; Leontaritis & Billings 1985; Morbidi et al. 2008; Schilling et al. 2001), only partial and quite weak results have been obtained in terms of time-varying function approximation and time-varying parameter estimation. Estimation of the states using artificial neural networks (ANN) has been presented in (Schilling et al. 2001).

The main contribution of this Chapter is the introduction of a new time-varying common-structured (TVCS) modelling scheme as a solution to the time-varying nonlinear systems identification problem, where the selection of the common model structure is the critical step throughout the modelling procedure. A new efficient CMSS algorithm is investigated to select a common model structure using an online sliding window approach. Once the common-structured model has been determined, relevant time-varying model parameters can then be estimated using a SWRLS algorithm. The novel study of common-structured model identification is particularly useful for engineering system design and control, where only a fixed common model structure is involved but with time-varying parameters.

The advantage of the proposed method is that first, even without a priori knowledge of the nonstationary system, a TVCS model can produce less biased or preferably an unbiased robust model with better generalisation properties. Second, the proposed model can be used to track fast and capture transient variations of varying parameters and can also be applied to study the performance of the behaviour of the underlying dynamical system characteristics. A simulated example and an application to real EEG data are included to demonstrate the performance of the new method.

5.2 The time-varying linear-in-the-parameter regression model

The identification problem of a nonlinear dynamical system is based on the observed input-output data $\{u(t), y(t)\}_{t=1}^{N}$, where $u(t)$ and $y(t)$ are the observations of the system input and output, respectively (Leontaritis & Billings 1985). This Chapter considers a class of discrete stochastic nonlinear systems which can be represented by the following nonlinear autoregressive with exogenous inputs (NARX) structure below (Chen et al. 2008; Wei & Billings 2009):

$$y(t) = f\left(y(t-1), \cdots, y(t-n_y), u(t-1), \cdots, u(t-n_u), \theta(t)\right) + e(t), \quad (5.1)$$

where $u(t)$ and $y(t)$ are the system input and output variables, respectively, n_u and n_y are the maximum input and output lags, respectively, $f(\cdot)$ is the unknown system mapping, and the observation noise $e(t)$ is an uncorrelated zero mean noise sequence providing that the function $f(\cdot)$ gives a sufficient description of the system. $X(t) = \left[y(t-1), \cdots, y(t-n_y), u(t-1), \cdots, u(t-n_u)\right]^T$ denotes the system 'input' vector with a known dimension $d = n_y + n_u$, and $\theta(t)$ is an unknown parameter vector. The NARX model (5.1) is a special case of the polynomial NARMAX model that takes the form below (Billings et al. 2007)

$$\begin{aligned} y(t) = f\{&y(t-1), \cdots, y(t-n_y), u(t-1), \cdots, \\ &u(t-n_u), e(t-1), \cdots, e(t-n_e); \theta(t)\} + e(t), \end{aligned} \quad (5.2)$$

The NARMAX model (5.2) was developed and discussed in (Chen & Billings 1989).

The non-linear mapping $f(\cdot)$ can be constructed using a class of local or global basis functions including radial basis functions (RBF), neural networks, multi-wavelets and different types of polynomials (Billings & Wei 2005b; Billings et al. 2007; Chen & Billings 1989; Chen et al. 2008; Leontaritis & Billings 1985; Li et al. 2011a; Wei et al. 2009). The polynomial model representation of a nonlinear

time-varying NARX is represented below

$$y(t)=\theta_0(t)+\sum_{i_1=1}^{d}\theta_{i_1}(t)x_{i_1}(t)+\sum_{i_1=1}^{d}\sum_{i_2=i_1}^{d}\theta_{i_1,i_2}(t)x_{i_1}(t)x_{i_2}(t)+$$

$$\cdots+\sum_{i_1=1}^{d}\cdots\sum_{i_d=i_{d-1}}^{d}\theta_{i_1,\cdots,i_d}(t)x_{i_1}(t)\cdots x_{i_d}(t)+e(t), \tag{5.3}$$

where $\theta_0(t)$, $\theta_{i_1,\cdots,i_m}(t)$ $(m=1,2,\cdots,d)$ are time-varying parameters and

$$x_k(t)=\begin{cases}y(t-k) & 1\le k\le n_y\\ u\left(t-\left(k-n_y\right)\right) & n_y+1\le k\le d\end{cases}, \tag{5.4}$$

The degree of a multivariate polynomial is defined as the highest order amongst the terms. If the number of regressors is d and the maximum polynomial degree is λ, the number of polynomial terms is $n_{\theta(t)}=\dfrac{(\lambda+d)!}{(\lambda!d!)}$. For large lags and the regression model (5.1) often involves a large number of candidate model terms, even if the nonlinear degree is not very high. For example, if $d=10$ and $\lambda=3$, then $n_{\theta(t)}=286$. Modelling experience has shown that an initial candidate model with a large number of candidate model terms can often be drastically reduced by including in the final model only the effectively selected significant model terms. The main motivation of this Chapter is to select significant common-structured model terms to form a parsimonious common model structure which generalises well (Aguirre & Billings 1995a).

A general form of the time-varying linear-in-the-parameter regression model is given as (Wei & Billings 2009)

$$y(t)=\sum_{m=1}^{M}\theta_m(t)\phi_m(t)+e(t)$$

$$=\varphi^T(t)\Theta(t)+e(t), \tag{5.5}$$

where M is the total number of candidate regressors. $\phi_m(t)=\phi_m\left(X(t)\right)$ $\left(m=1,\cdots,M\right)$ are nonlinear functions and $\theta_m(t)$ $\left(m=1,\cdots,M\right)$ represents the

model time-varying parameters. $\varphi(t)=\left[\phi_1\left(X(t)\right),\cdots,\phi_M\left(X(t)\right)\right]^T$ and $\Theta(t)$ are the associated regressor and parameter vectors, respectively. It should be noted that in most cases the initial full regression Eq. (5.5) might be highly redundant. Some of the regressors or model terms can be removed from the initial regression equation without any effect on the predictive capability of the model, and this elimination of the redundant regressors usually improves the model performance (Aguirre & Billings 1995a). For most nonlinear dynamical system identification problems, only a relatively small number of model terms are commonly required in the regression model. Thus an efficient model term selection algorithm is highly desirable to detect and select the most significant regressors.

5.3 TVCS model identification

The CMSS algorithm is a critical step in TVCS identification. Once the common-structured model has been identified, relevant model parameters for each window data set can then be estimated, and the transient properties of the model parameters on the associated data set can thus be deduced. The identification procedure for TVCS models contains the following steps.

5.3.1 Data acquisition

For an original N-sample observational input-output data $D_N=\left\{u(t),y(t)\right\}_{t=1}^{N}$, the $K+1$ datasets can be obtained by using an online sliding window size W, with 50% overlap, where the parameter $K+1$ is equal to $\lceil N/(W/2)\rceil-1$ and $\lceil x\rceil$ denotes taking the upper integer part of the variable x. Note that how to choose the suitable choice of window size W is discussed in (Hwang et al. 2002).

5.3.2 The common model structure selection (CMSS) algorithm

Assume that a total of $(K+1)$ data sets, where the first K represent the training data sets, and the last data set is used as a test data set, obtained by the online sliding-window have been carried out on the same system. Also, assume that a common model structure of Eq. (5.5) can be best fit to all the training data sets. Denote the observed input-output sequences for the k th data set by $D_{N_k} = \{u_k(t), y_k(t)\}_{t=1}^{N_k}$ for $k = 1, \cdots, K+1$. Thus the k th 'input' vector is represented by $X_k(t) = [x_{k,1}(t), \cdots, x_{k,d}(t)]^T = [y_k(t-1), \cdots, y_k(t-n_y), u_k(t-1), \cdots, u_k(t-n_u)]^T$. Assume that all the K training data sets can be represented using a common model structure for the different parameters, then the initial candidate multiple regression model can be formulated as (Wei & Billings 2009; Wei et al. 2008)

$$\begin{aligned} y_k(t) &= \sum_{m=1}^{M} \theta_{k,m} \phi_m(X_k(t)) + e_k(t) \\ &= \sum_{m=1}^{M} \theta_{k,m} \phi_{k,m}(t) + e_k(t), \end{aligned} \tag{5.6}$$

where the parameters $\theta_{k,m}$ in Eq. (5.6) are time-independent constants, Eq. (5.6) will be called the time-invariant common structured model. If the parameters $\theta_{k,m}$ are time-dependent, the time-varying common structure (TVCS) model is represented by

$$\begin{aligned} y_k(t) &= \sum_{m=1}^{M} \theta_{k,m}(t) \phi_m(X_k(t)) + e_k(t) \\ &= \sum_{m=1}^{M} \theta_{k,m}(t) \phi_{k,m}(t) + e_k(t), \end{aligned} \tag{5.7}$$

where $\phi_{k,m}(t) = \phi_m(X_k(t))$ for $k = 1, 2, \cdots, K$, $m = 1, 2, \cdots, M$, and $t = 1, 2, \cdots$

, N_k. The representation of Eq. (5.7) using a compact matrix form can be expressed as

$$\Upsilon_k = \Phi_k \Theta_k + \mathrm{E}_k, \tag{5.8}$$

where $\Upsilon_k = \left[y_k(1), \cdots, y_k(N_k) \right]^T$, $\Theta_k = \left[\theta_{k,1}(t), \cdots, \theta_{k,M}(t) \right]^T$, $\mathrm{E}_k = [e_k(1), \cdots, e_k(N_k)]^T$, and $\Phi_k = \left[\varphi_{k,1}, \cdots, \varphi_{k,M} \right]$ with $\varphi_{k,m} = \left[\phi_{k,m}(1), \cdots, \phi_{k,m}(N_k) \right]^T$, for $k = 1, 2, \cdots, K$ and $m = 1, 2, \cdots, M$.

A new CMSS algorithm will be developed to select a common-structured sparse model from the multiple regressions shown in Eq. (5.6) and (5.7). Let $I = \{1, 2, \cdots, M\}$ and denote $D = \{\phi_m : m \in I\}$ as the dictionary of candidate model terms. For the k th window data set, the dictionary D can be used to form a dual dictionary $\Psi_k = \{\varphi_{k,m} : m \in I\}$ note that the m th candidate basis vector $\varphi_{k,m}$ is formed by the m th candidate model term $\phi_m \in D$. Thus the CMSS problem is equivalent to finding a subset $\{\phi_{p_1}, \phi_{p_2}, \cdots, \phi_{p_n}\} \subset D$ (normally $n \le M$). So that Υ_k $(k = 1, 2, \cdots, K)$ can be approximated using a linear combination of regression terms $\{\varphi_{k,p_1}, \cdots, \varphi_{k,p_n}\} \subset \Psi_k$ below

$$\Upsilon_k = \theta_{k,1}(t)\varphi_{k,p_1} + \cdots + \theta_{k,n}(t)\varphi_{k,p_n} + \mathrm{E}_k, \tag{5.9}$$

The next step of the CMSS algorithm selects significant model terms in a forward stepwise way. The first significant common model term can be selected as the p_1 th element, $\phi_{p_1} \in D$ by maximising the sum of error reduction ration (ERR) (Wei et al. 2008) values for all the K data sets from I. Thus the first significant basis vector for the k th regression model is $\alpha_{k,1} = \varphi_{k,p_1}$, and the associated orthogonal basis vector can be chosen as $q_{k,1} = \varphi_{k,p_1}$. Generally, the m th significant model term of k th regression model ϕ_{k,p_m} can be chosen by the following steps. It is

assumed that at the $(m-1)$th step, $(m-1)$ significant model terms, namely, $\{\phi_{k,1},\cdots,\phi_{k,m-1}\}$, have been selected by maximising the $(m-1)$th average ERR (AERR) for all the K data sets from I, which guarantees that the variation of the outputs in all the K data sets with the highest percentage, compared with choosing any other candidate model term $\phi \in D$. The AERR criterion provides a way to select significant vectors one by one. Once the first $(m-1)$ basis vectors $\{\alpha_{k,1},\cdots,\alpha_{k,m-1}\}$ have been determined, and the associate orthogonal vectors $\{q_{k,1},\cdots,q_{k,m-1}\}$ can be obtained, then these $(m-1)$ vectors together with the m th vector $\alpha_{k,m}=\varphi_{k,p_m}$, and the associated orthogonal vector $q_{k,m}$, can be selected step-by-step. Further details and derivation of the procedure of the CMSS algorithm can be found in (Wei et al. 2008). It should be noted that, in this Chapter, a time varying common structure model as a solution to time varying nonlinear system identification problems can be obtained with the CMSS algorithm using an online sliding window approach to track the variations and capture the transient information for the EEG signals, which is totally different from the CMSS approach based on several data sets collected from different experiments discussed in (Wei et al. 2008).

To determine the proper common model size, the generalized cross-validation (GCV) criterion (Billings et al. 2007) can be adopted to terminate the CMSS procedure. Specially, for the l-term model, the GCV of single regression model is defined as $GCV(l)=\left(\frac{N}{N-\mu_m l}\right)^2 MSE(l)$, where MSE is the mean squared error, $\mu_m=\max\{1,\rho N\}$ and $0\le\rho\le 0.01$. As a rule of thumb, a good choice for μ_m is to use a value from the range of $5\le\mu_m\le 10$. The average GCV (AGCV) is formulated by

$$AGCV(l)=\frac{1}{K}\sum_{k=1}^{K}GCV^{[k]}(l), \tag{5.10}$$

where $GCV^{[k]}(l)$ is the value for the GCV criterion associated to the k th data set. If the AGCV reaches the minimum at $l=n$, then the CMSS procedure is terminated, yielding an n -term model.

5.3.3 Model parameter estimation

The parameters for the common structured model (5.6) can be easily calculated by

$$A_{k,n}=Q_{k,n}R_{k,n}, \tag{5.11}$$

where $A_k=\left[\varphi_{k,p1},\cdots,\varphi_{k,p_n}\right]$, $Q_{k,n}$ is an $N_k\times n$ matrix with orthogonal columns $q_{k,1},q_{k,2},\cdots,q_{k,n}$, and $R_{k,n}$ is an $n\times n$ unit upper triangular matrix whose entries are calculated during the orthogonalisation procedure. For TVCS model of Eq. (5.7), it is also easy to calculate the value of the unknown time-dependent parameters by SWRLS algorithm (Choi & Bien 1989) for each data window of the $(K+1)$th data sets. A sliding window adaptation algorithm is similar to the RLS algorithm in terms of both the derivation and convergence characteristics. Here the SWRLS algorithm is applied to estimate the time varying parameters in the model (5.7) that to be able to more easily track nonstationary processes. The finite sliding window allows for any data outliers to be forgotten after a finite number of iterations. The sliding window RLS algorithm provides much increased convergence rate at the expense of increased computational complexity. The transient variation properties of the observational data can thus be deduced by the transient parameter values for the associated data set.

5.4 Case study

Two examples are provided to illustrate the performance of the proposed TVCS model identification procedure. The data used in the first example are simulated from a known nonstationary model; this is a severely nonstationary process. The objective here is to illustrate the capability of novel TVCS approach for tracking and capturing the transient variation for the time-varying parameters. The second example involves a practical modelling problem of EEG data.

5.4.1 Example 1: Simulation data

Prior to applying the proposed TVCS modelling approach to real EEG data, an artificial time-varying signal was considered. The signal below was simulated

$$y(t)=\sum_{i=1}^{7} A_i \cos\left(2\pi f_i t+\varphi_0\right), \tag{5.12}$$

where $A=\dfrac{1}{(2k-1)}$ with $k=1,2,\cdots,7,$ $f=[50,150,250,350,450,600,750]$, initial phase shift $\varphi_0=\dfrac{3\pi}{2}$, and sample time t is 0.08 second, respectively. The above signal was sampled with a sampling interval 0.0001, and thus a total of 800 observations were obtained. A Gaussian white noise sequence, with mean zero and variance of 0.04, was then added to the 800 data points.

The objective is to identify a TVCS model, and then the transient dynamical properties of the analytical signal can be deduced from the time-varying parameters. Denote the system output sequences using $\{y(t)\}_{t=1}^{N}$, with N = 800. The sliding window size W = 200 is applied to obtain the $K+1=N/(W/2)-1$ = 7 data sets. Here from the properties of the simulation signal，the sliding window length should be chosen as W = 200, with 50% overlap. First 6 training data sets were used for the common-structured model identification, and the 7th data set was used to test the performance of the identified model. The predictor vector for all the

common-structured models was chosen to be $X(t)=\left[x_1(t),\cdots,x_5(t)\right]^T$, where $x_k(t)=y(t-k)$ for $k=1,\cdots,5$. The initial common structure for all the first 6 training data sets was chosen to be a NAR model below

$$y(t)=\theta_0+\sum_{i=1}^{5}\theta_i x_i(t)+\sum_{i=1}^{5}\sum_{j=i}^{5}\theta_{i,j}x_i(t)x_j(t)+e(t), \tag{5.13}$$

This candidate model involves a total of 21 candidate model terms. Based on the candidate common model structure, the novel CMSS algorithm was applied to the first 6 training data sets. The AGCV criterion, shown in Figure 5.1, suggests that a common model structure, with 6 model terms, is preferred. The 6 selected common model terms, ranked in order of significance are shown in Table 5.1. Now consider the performance of the identified model, whose parameters are determined by Eq. (5.11) and Table 5.1. The 7th test data set, which has never been used in the identification procedure, was applied to test the performance of the identified model. Figure 5.2 presents a comparison between the recovered signal from the identified common structured model and the original measurements.

To measure the identified models, the normalized root-mean-squared-error (RMSE) is defined as follows:

$$RMSE=\sqrt{\frac{1}{N_{K+1}}\sum_{t=1}^{N_{K+1}}\left\|\frac{\hat{y}(t)-y(t)}{y(t)}\right\|^2}, \tag{5.14}$$

where N_{K+1} is the data sliding window length of the $(K+1)$th test data set, $\hat{y}(t)$ is the predicted value from the identified model. The RMSE criteria in Eq. (5.14) can also be provided to select a proper sliding window size W provided that the RMSE value is very small. The value for RMSE, for the identified models, over the test window data set, was calculated as 2.2051. Clearly identified model provides an excellent presentation for the test data set.

The TVCS model was thus represented by

$$y(t)=\theta_0(t)+\sum_{i=1}^{4}\theta_i(t)y(t-i)+\theta_5 y(t-2)\,y(t-3)+e(t), \qquad (5.15)$$

where the parameter $\theta(t)$ depends on the data sets from the sliding window. The parameters can be directly estimated using the SWRLS algorithm. Figure 5.3 shows the estimated values for $\theta(t)$ for the test data set given in Figure 5.2 using the SWRLS algorithm with a forgetting factor of 0.998. The time-varying coefficients estimates in Figure 5.3 can give more transient information, for example, there are two clear abrupt changes of the estimated coefficients at sample index interval from 60 to 80, and from 100 to 120, respectively, which show that the original signal shown in Figure 5.2 undergoes transient changes. Furthermore, the proposed method can also track and detect variation of each training data block dynamically, for example, Figure 5.5 shows the rapid change of coefficient estimation at sample index about from 60 to 65 and about 100, respectively, which implies that the corresponding original training data block given in Figure 5.4 changed at sample index from 60 to 65 and about 100, respectively. The results discussed are shown that the CMSS algorithm is effective.

Table 5.1 Identification results for the simulation data with the CMSS algorithm for NAR model representation

Step	model term	Parameters for test data sets						AERR (%)
		Data01	Data 02	Data 03	Data 04	Data 05	Data 06	
1	$y(t-1)$	0.4662	0.4884	0.4401	0.4930	0.4304	0.5612	88.3984
2	$y(t-3)$	0.3078	0.2663	0.2249	0.2770	0.3887	0.2567	1.5840
3	$y(t-2)$	0.2279	0.1766	0.1830	0.2274	0.3187	0.2176	0.4903
4	$y(t-4)$	0.2096	0.2385	0.1769	0.0665	0.0159	0.0406	0.0667
5	*Const*	-0.0790	0.0224	-0.1046	0.0075	-0.0586	0.0108	0.0118
6	$y(t-2)y(t-3)$	0.1098	-0.0246	0.1577	-0.0033	0.0803	-0.0053	0.0826

RMSE = 2.205

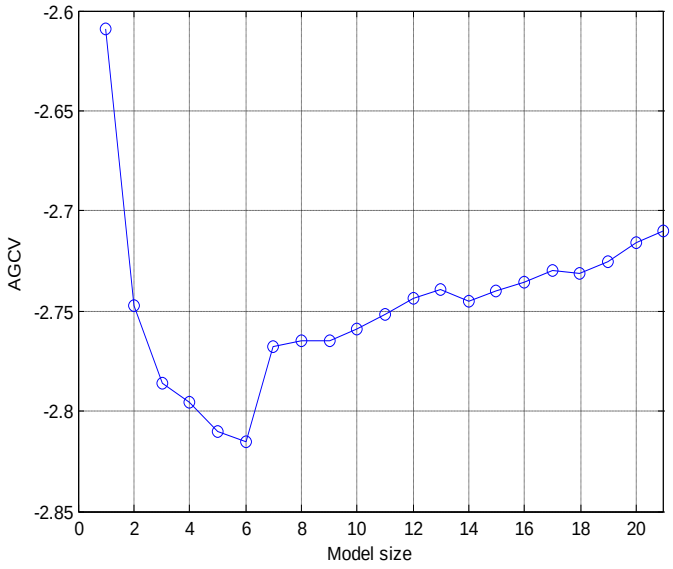


Figure 5.1 AGCA versus model size for common model structure selection models over the output signals.

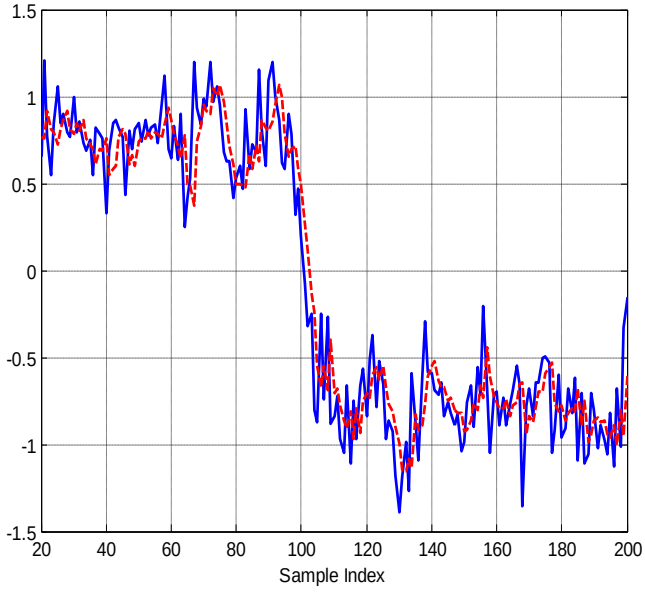


Figure 5.2 A comparison of the recovered signal from the identified TVCS model (5.15) and the original observations. Solid (blue) line indicates the observations and the dashed line indicates the signal recovered from the TVCS model (5.15).

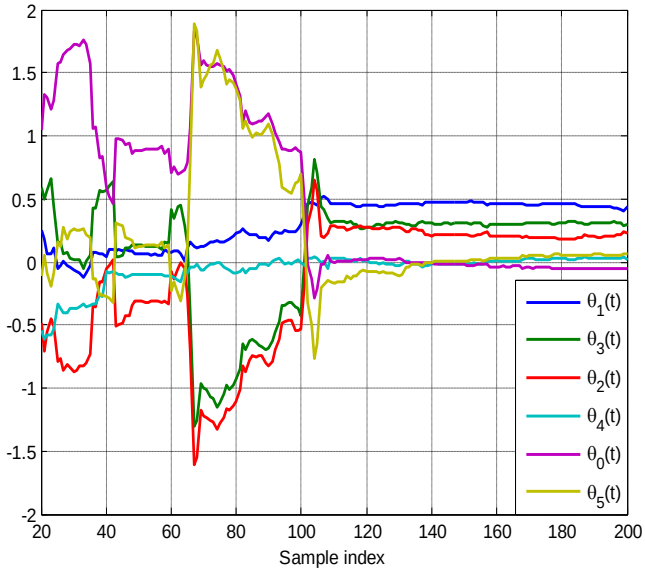


Figure 5.3 The time-varying coefficients estimation for test data set of NAR identified Common- structured model in Eq. (5.15) using the SWRLS algorithm with a forgetting factor of 0.998.

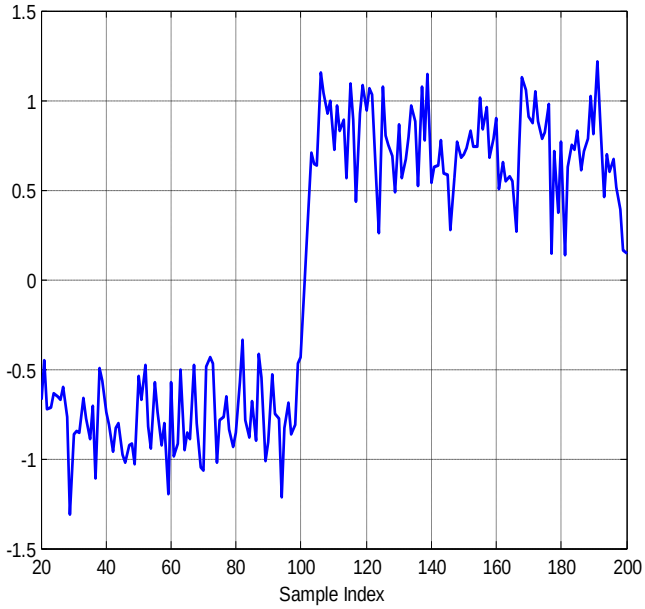


Figure 5.4 The simulation training data block output for the 4th window block data set.

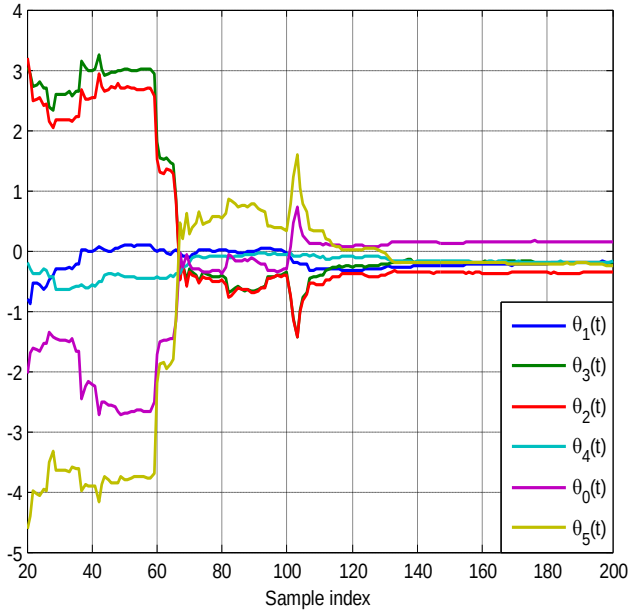


Figure 5.5 The time-varying coefficients estimation of NAR identified common-structured model in Eq. (5.15) for training data set given in Figure 5.4 using the SWRLS algorithm with a forgetting factor of 0.998.

5.4.2 Example 2: modelling EEG data

5.4.2.1 EEG datasets

EEG series provides an illustrative analysis to highlight key features of the methodology. Scalp EEG series are synchronous discharges from cerebral neurons detected by electrodes attached to the scalp. An XLTEK 32 channel headbox (Excel-Tech Ltd) with the international 10-20 electrode placement system was applied to record EEG data in the Sheffield Teaching Hospitals NHS Foundation Trust, Royal Hallamshire Hospital, UK. 32 parallel EEG series were recorded in parallel from 32 electrodes located on epileptic seizure patient's scalp using the same-32 channel amplifier system using bipolar montages reference channels. The

sampling frequency of the device was 500 Hz. Dynamical properties of brain electrical activity from different extracranial and intracranial recording regions has been discussed in (Andrzejak et al. 2001). The time-frequency decomposition method aided by the time-varying autoregressive (TVAR) model for EEG series to extract and estimate latent EEG components in various key frequency bands was also investigated in (Li et al. 2011b; West et al. 1999). The central objective of this paper for the EEG signals is to propose an empirical and data-based TVCS modelling scheme to track and capture the transient variations of EEG signals from model identification that can produce an accurate but simple description of the dynamical relationships between different recording regions during brain activity. This is a complicated and challenging black box system where the true model structure is unknown, and thus, needs to be identified from the available experimental data. As an example, symmetrical two bipolar channels (F3, located over the left superior frontal area of the brain and F4, located over the same area on the right) of EEG recorded from a patient with absence seizure epileptic discharge is investigated. Channel F3 was treated as the input, denoted by $u(t)$, and Channel F4 was treated as the output, denoted by $y(t)$, note that Channel F3 is the signal input and Channel F4 is the signal output, the main reason is that the phase of Channel F4 is related to the phase of Channel F3. The objective is to investigate, from the available Channel F3 and F4 recordings, if an identified TVCS model is suitable to describe the dynamical characteristics and adaptively track and capture the transient variations of time-varying parameters using the proposed approach. The input-output EEG signals of N = 3000 data points pairs of one seizure, which are for a sort of epileptic seizure activity of a patient, with a sampling rate of 500 Hz, recording during 6 seconds, were analysed in this example. This analysis represents the first application of TVCS modelling to epileptic seizure EEG data and was intended to determine feasibility and identify its potential for tracking transient variations over time in seizure EEG data.

5.4.2.2 TVCS Model identification

Similar to the previous simulation example, the objective is to identify a TVCS model which can be used to analyse and detect transient variation properties of EEG signals and dynamically track and capture the variation of the EEG signals. Simulation results have shown that, the choice of sliding window size W = 600 data points, gives good model identified results. So the parameter $K+1$ was equal to 9. The first 8 datasets will be considered as training data sets for the model identification, and the 9th test data set which has never been used in the identification procedure was then used to test the performance of the identified model. Denote the system input and output sequence using $D_N = \{u(t), y(t)\}_{t=1}^{N}$ with N = 3000 data pairs. The predictor vector for all the common-structured models was chosen to be $X(t) = [x_1(t), \cdots, x_{10}(t)]^T$, where $x_k(t) = y(t-k)$ for $k = 1, 2, \cdots, 5$ and $x_k(t) = u(t-k+5)$ for $k = 6, 7, \cdots, 10$. The initial candidate common model structure for all the 8 training data sets was chosen to be a NARX model below

$$y(t) = \theta_0 + \sum_{i=1}^{10} \theta_i x_i(t) + \sum_{i=1}^{10} \sum_{j=i}^{10} \theta_{i,j} x_i(t) x_j(t) + e(t), \tag{5.16}$$

This candidate model involves a total of 66 candidate model terms. Based on the candidate common model structure, the new CMSS algorithm was applied to the 8 training data sets to identify a TVCS model. The AGCV index, shown in Figure 5.6, suggests that a common model structure, with 8 model terms is preferred. The 8 selected common model terms, ranked in order of the significance, are shown in Table 2.2. The common model structure for the 8 training data sets was identified as

$$\begin{aligned} y(t) = \theta_0 + \sum_{i=1}^{4} \theta_i y(t-i) + \theta_5 u(t-1) + \theta_6 y(t-5) \\ \times u(t-1) + \theta_7 y(t-5) u(t-5) + e(t), \end{aligned} \tag{5.17}$$

The corresponding coefficient parameters in (5.17) for each training window block data set are given in Table 5.2. Figure 5.7 presents a comparison between the recovered signal and the associated measurements, where the relevant normalized RMSE, with respect to the test data set, was calculated to be 0.2755. Clearly, the TVCS model can provide an excellent representation for the test data set.

The common structured model (5.17) was then employed to form a TVCS model

$$y(t)=\theta_0(t)+\sum_{i=1}^{4}\theta_i(t)y(t-i)+\theta_5(t)u(t-1)+\theta_6(t)$$

$$\times y(t-5)u(t-1)+\theta_7(t)y(t-5)u(t-5)+e(t), \tag{5.18}$$

where $\theta_i(t)$ $(i=0,2,\cdots,7)$ are now time-dependent parameters which can then be estimated by using the SWRLS algorithm. The associated parameter estimates are shown in Figure 5.8. The time-varying model estimation results here, in combination with an extension of the concept of the generalized frequency response functions (GFRFs) (Billings & Jones 1990; Jones & Billings 1989), can be used to form some nonlinear parametric time-frequency formulas, which can then be used to generate nonlinear time-frequency properties that are useful for feature extraction from EEG data.

Note that polynomial models may be intrinsically unstable in some cases (Ozaki et al. 1999) if a full model is directly used in a simple manner. The proposed approach would be to only select the appropriate polynomial terms which then avoid the unstable problems. This is why term selection becomes so important in nonlinear system identification.

Table 5.2 Identification results for the EEG data with the CMSS algorithm for NARX model representation

Step	Model term	Parameters for the EEG data sets								AERR (%)
		Data01	Data 02	Data 03	Data 04	Data 05	Data 06	Data 07	Data 08	
1	$y(t-1)$	1.9406	1.8059	1.8513	1.7979	1.4649	1.2945	1.3323	1.7458	97.1551
2	$y(t-2)$	-1.3804	-1.2920	-1.3627	-1.2635	-0.7640	-0.4772	-0.5121	-1.1687	1.0140
3	$y(t-3)$	0.6155	0.6372	0.6665	0.5496	0.3049	0.2344	0.2097	0.5779	0.0688
4	$y(t-4)$	-0.2304	-0.2151	-0.2299	-0.1315	-0.0542	-0.0761	-0.0250	-0.1549	0.0381
5	$y(t-5)u(t-1)$	-0.0001	-0.0001	-0.0001	-0.0002	-0.0003	-0.0003	-0.0005	-0.0003	0.0149
6	*Const*	-2.9959	-6.7214	-9.5099	-5.8357	-5.3187	-2.9343	1.6203	-0.3837	0.0138
7	$u(t-1)$	0.0283	0.0344	0.0433	0.0291	0.0288	0.0079	-0.0164	-0.0049	0.0371
8	$y(t-5)u(t-5)$	0.0001	0.0001	0.0001	0.0002	0.0004	0.0004	0.0005	0.0004	0.0351

RMSE is 0.2755%.

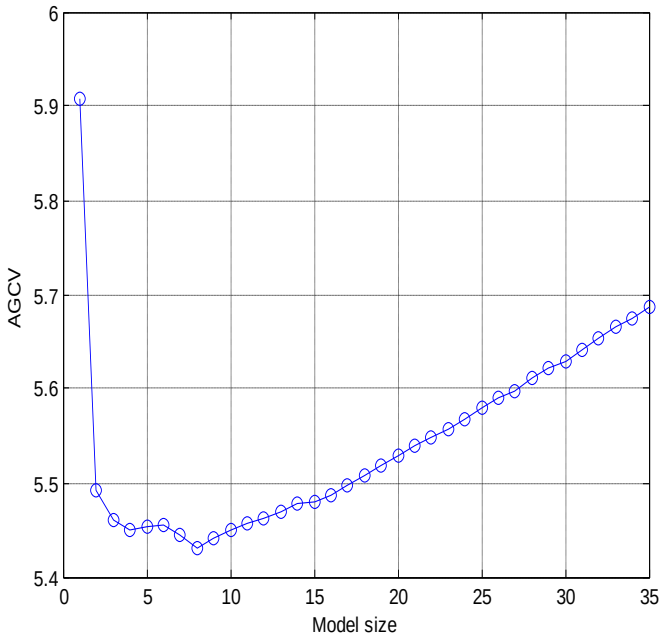


Figure 5.6 AGCV versus model size for common model structure models of the input-output EEG signals.

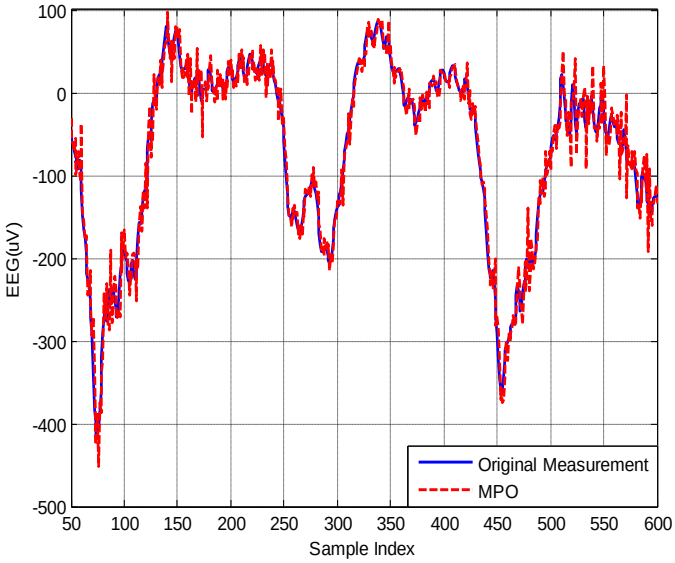


Figure 5.7 A comparison of the recovered signal from the identified TVCS model (5.18) and the original observations over the test data set. Solid (blue) line indicates the observations and the dashed line indicates the signal recovered from the TVCS model.

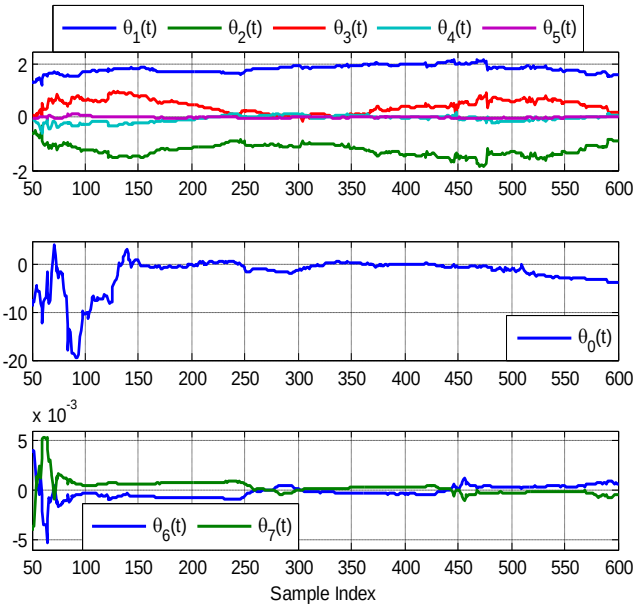


Figure 5.8 The time-varying coefficient estimates of nonlinear common-constructed model Eq. (5.18) for the 9th EEG test data set with a SWRLS algorithm with a forgetting factor of 0.98.

5.5 Conclusions

Time varying nonlinear system identification using the new CMSS algorithm and online sliding-window approach was studied in this chapter. The application of the new CMSS approach involves two critical steps: model structure selection and model parameter estimation. When the CMSS algorithm is applied in model structure selection, a multiple regression search procedure, over a number of partitioned data sets using an online sliding-window approach, is performed. Initially the implementation of a multiple search appears to be very complex. But the introduction of the new multiple orthogonal regression search algorithm provides an attractive solution to this problem. It should be noted that the computational complexity of the CMSS algorithm depends on the block data sets K, where the parameter K depends on the sampled data length N and the sliding window size W. The choice of the sliding window size W depends on the properties of the observational data. The true model structure of the underlying system will in many cases be unknown and only the input and output observations are available. But the algorithms derived in this Chapter show that a common model structure can be deduced from the available observations. In the two examples, polynomial models were employed to form the common-structured models. However, it should be noted that the CMSS approach can also apply to any other parametric or non-parametric modelling problems where the initial full models can be written as a linear-in-the- parameters form.

Once a common model structure has been obtained, an online sliding-windowing recursive least squares (SWRLS) algorithm is then applied to estimate the time-varying model parameters. But note that other online methods for example a sliding-window Kalman filtering algorithm can also be employed to estimate the unknown time-varying parameters. While the tracking ability of RLS is achieved by performing a forgetting operation on the information matrix, the tracking

capability of Kalman filtering is obtained by adding a nonnegative definite matrix to the covariance matrix. The main reason for employing RLS in the present Chapter is mainly because of its simple calculation and good convergence properties, but this does not exclude the introduction of a Kalman filtering algorithm in our future comparative studies.

The TVCS model can be applied to analyze and reflect the transient properties of nonstationary signals, and also to dynamically track and capture the transient variation of the nonstationary EEG signals. The main purpose of this Chapter at this stage is focused on nonlinear time-varying parametric modelling, which forms the basis of some important developments for further application in medical applications including EEG data modelling, analysis, and feature extraction. For example, the time-varying parameter results estimated from the TVCS model cannot only provide the transient local information of the EEG signals, but can also be applied in nonlinear time-dependent parametric spectral analysis in the frequency domain to extract more features from the EEG signals, so that the results can provide further applications for EEG data analysis.

In the next Chapter, the time-varying linear and nonlinear Granger causality will be investigated using the NARMAX approach to analyse the directed connections between EEG channels.

Chapter 6

Time-Varying Linear and Non-linear Parametric Models for Granger Causality Analysis

6.1 Introduction

This chapter aims to establish a time-varying cause-effect modelling approach based upon the NARMAX model identification techniques, with application to EEG data to detect the directed causal influence between EEG channels.

In cognitive neuroscience, as in many other science and engineering research fields, the investigation of EEG data is usually carried out by using measures of correlation, coherence and mutual information (Quiroga et al. 2002). These measures, however, provide little insight into the directionality of information flow. A question of great interest is whether there exist causal relations among a set of measured variables. Several recent works based on vector autoregressive (VAR) models have begun to consider this problem (Verdes 2005). Causal relations between different components of a multi-dimensional signal can be analysed in the context of multivariate autoregressive modelling. The most popular approach to deal with the causal relations is the so-called Granger causality method (Granger 1969). The major approach to causality examines if the prediction of one series could be improved by incorporating information of the other, as discussed by Granger. Particularly, if the prediction error of the signal X is reduced by including measurements from the signal Y in the regressor model, then the signal Y is said to have a causal influence on the signal X. Granger causality was originally investigated for linear systems (Granger 1969). Recently this concept has

been extended to the nonlinear case. The application of Granger causality to neuroscience data has been discussed for *f*MRI (Sato et al. 2006), EEG (Blinowska et al. 2004; Hesse et al. 2003) and MEG experiments (Gow et al. 2008).

Linear time-varying causality was previously investigated on scalp EEG (Hesse et al. 2003). Hesse et al. (2003) has studied the linear recursive time-variant estimation of the Granger causality based on the adaptive recursive fit of a VAR model with time-varying parameters using a recursive least squares (RLS) algorithm, where the assumption of stationarity of the signals can be removed. Recently the Granger causality definition has been extended to nonlinear bivariate time series (Ancona et al. 2004; Marinazzo et al. 2006).

All traditional Granger causality detection methods are based on the time-invariant linear ARX models or time-invariant nonlinear models. It follows that standard linear VAR models may not always be able to capture the dynamic behavior of many nonstationary time series. To the best of our knowledge, results on time-varying nonlinear Granger causality analysis have seldom been reported in the literature. In this Chapter, we will introduce a novel nonlinear method that can be used to detect and track nonlinear dynamical Granger causalities. A delay vector variance (DVV) method based on the examination of local predictability (Mandic et al. 2008) is first applied to detect the presence of nonlinearity for EEG epileptic data. We will then use time-varying Granger causality to obtain dynamic characterizaiton of causal interaction estimates among cortical areas based on the adaptive recursive fit of a VAR model with time-varying parameters using a recursive least squares algorithm. To illustrate the performance of the new method, several examples are presented: one for artificial data where the exact causal effect feature is known, and another two for real EEG data sets where the hidden causality feature is revealed by the new method.

6.2 Method

6.2.1 Time-varying linear Granger causality

The Granger causality is a fundamental tool for the description of causal interactions of two time series. We detail the bivariate case of the Granger causality in this Chapter.

6.2.1.1 Time-invariant Granger causality

Let X and Y be the two signals whose time observations are noted $x(t)$ and $y(t)$ with $t=1,2,\cdots,N$. In order to show the improvement of the prediction of one signal by taking into consideration the past of the second signal, the univariate AR and bivariate ARX models are fitted to the signals, respectively. If the temporal dynamics of $x(t)$ and $y(t)$ are suitably represented by a time-invariant univariate AR model of order p, we can obtain

$$x(t)=\sum_{i=1}^{p} a_{1,i}x(t-i)+u_1(t), \tag{6.1}$$

$$y(t)=\sum_{i=1}^{p} b_{1,i}y(t-i)+u_2(t), \tag{6.2}$$

where the model prediction error u_1 and u_2 depend only on the past of the own signal. The time-invariant bivariate ARX(p,q) models are represented by

$$x(t)=\sum_{i=1}^{p} a_{2,i}x(t-i)+\sum_{l=1}^{q} c_{2,l}y(t-l)+v_1(t), \tag{6.3}$$

$$y(t)=\sum_{i=1}^{p} b_{2,i}y(t-i)+\sum_{l=1}^{q} d_{2,l}x(t-l)+v_2(t), \tag{6.4}$$

where the prediction error v_1 and v_2 depend on the past of the signal itself and additionally on the past of the second signal. The coefficients in the model (6.1)-(6.4) are generally estimated by solving the Yule-Walker equations (Proakis & Manolakis 1996; Subasi 2007) which require the stationarity of the signals and result in a time-invariant VAR model analysed over the time course.

Let us begin with the bivariate case of causality $X \to Y$. The reciprocal case is similar. The accuracy of prediction in model (6.1) and (6.2) may be evaluated by the unbiased variance of the prediction errors $\sum_{y|y^-}$ where y^- symbolizes y past

$$\sum\nolimits_{y|y^-} = \frac{1}{N-p}\sum_{t=1}^{N} u_2^2(t) = \frac{RSS_{y|y^-}}{N-p} = \mathrm{var}(u_2), \tag{6.5}$$

where $RSS_{y|y^-}$ is the residual sum of squares in the model (6.2). For the bivariate model (6.3) and (6.4), we can obtain

$$\sum\nolimits_{y|y^-,x^-} = \frac{1}{N-p-q}\sum_{t=1}^{N} v_2^2(t) = \frac{RSS_{y|y^-,x^-}}{N-p-q} = \mathrm{var}(v_2), \tag{6.6}$$

If the signal X causes the signal Y in the Granger sense, the variance of the prediction error $\sum_{y|y^-,x^-}$ must be smaller than prediction error $\sum_{y|y^-}$. The linear Granger causality $X \to Y$ is then defined by (Gourevitch et al. 2006)

$$LGC_{x\to y} = \ln\frac{\sum_{y|y^-}}{\sum_{y|y^-,x^-}}, \tag{6.7}$$

Correspondingly, the linear Grange causality of $Y \to X$ is evaluated by

$$LGC_{y\to x} = \ln\frac{\sum_{x|x^-}}{\sum_{x|x^-,y^-}}, \tag{6.8}$$

Generally, the most important property of the Granger causality is the positivity, when a signal X causes a second signal Y. The Eq. (6.7) and (6.8) represent a simple measure for the strength of directional interaction.

6.2.1.2 Time-varying Causality measure

The time-varying fit of a VAR model is required to detect the transient directed interactions. Ding et al. (2000) investigated a VAR model fitting algorithm to obtain the time-varying Granger causality which requires to assume that the the signals to be studied are stationary within a short-time window, and the changes fro one window to another is smooth. There is a limitation in the effectiveness for this approach. First, if the processes are varying rapidly, a process assumed to be stationary may be too small to allow for sufficient accuracy in the estimation of the relevant parameters over the window. Second, this approach would not easily accommodate the step changes with the analysis intervals. Third, this solution imposes an incorrect model on the observal data, that is, piecewise stationary. Therefore, An adaptive recursive fit of a VAR model with time-dependent parameters by means of some adaptive filtering procedures such as recursive least squares (RLS), least mean squares (LMS) and Kalman filtering algorithms is proposed to capture the transient Granger causality. The time-varing VAR model fitting can yield time-varying autoregressive parameters. Consequently, by contrast with the model (6.1)-(6.4), the time-varying VAR models are represented by

$$x(t)=\sum_{i=1}^{p} a_{1,i}(t)x(t-i)+u_1(t), \tag{6.9}$$

$$y(t)=\sum_{i=1}^{p} b_{1,i}(t)y(t-i)+u_2(t), \tag{6.10}$$

and

$$x(t)=\sum_{i=1}^{p} a_{2,i}(t)x(t-i)+\sum_{l=1}^{q} c_{2,l}(t)\,y(t-l)+v_1(t), \tag{6.11}$$

$$y(t)=\sum_{i=1}^{p} b_{2,i}(t)y(t-i)+\sum_{l=1}^{q} d_{2,l}(t)\,x(t-l)+v_2(t), \tag{6.12}$$

The time-varying fit of VAR models yields time-varying variance of prediction error. A general recursive variance computational formula can be defined by

$$\sigma^2(t+1) = (1-c)\sigma^2(t) + c\Delta^2(t), \tag{6.13}$$

where the constant lies at $0 < c < 1$, and $\Delta(t)$ is one of $u_1(t)$, $u_2(t)$, $v_1(t)$, or $v_2(t)$ which represent the time-varying variances of the corresponding prediction errors $\sum_{x|x^-}(t)$, $\sum_{y|y^-}(t)$, $\sum_{x|x^-,y^-}(t)$, and $\sum_{y|y^-,x^-}(t)$.

Therefore, the representation of time-varying linear Granger causality is then evaluated as

$$LGC_{x\to y}(t) = \ln \frac{\sum_{y|y^-}(t)}{\sum_{y|y^-,x^-}(t)}, \tag{6.14}$$

$$LGC_{y\to x}(t) = \ln \frac{\sum_{x|x^-}(t)}{\sum_{x|x^-,y^-}(t)}, \tag{6.15}$$

The calculation of the time varying Granger causalities in Eq. (6.14) and (6.15) is analogous to Eq. (6.7) and (6.8). The time varying strength of interaction may be quantified by the maximum at each time point from Eq. (6.14) and (6.15).

6.2.2 Time-varying nonlinear Granger causality

In this subsection, our main purpose is to find the general VAR models suitable to evaluate Granger causality, thus extending the RBF model results discussed in (Marinazzo et al. 2006).

6.2.2.1 NARX model

The identification problem of a nonlinear dynamical system is based on the observed input-output data $\{x(t), y(t)\}_{t=1}^{N}$, where $x(t)$ and $y(t)$ are the observations of the system input and output, respectively (Leontaritis & Billings 1987). This Chapter considers a class of discrete stochastic nonlinear systems

which can be represented by the following nonlinear autoregressive with exogenous inputs (NARX) structure below (Chen et al. 2008; Leontaritis & Billings 1985; Ljung 2001; Wei & Billings 2004):

$$y(t)=f\left(y(t-1),\cdots,y(t-n_y),x(t-1),\cdots,x(t-n_x),\theta\right)+e(t), \quad (6.16)$$

where $f(\cdot)$ is the unknown typically nonlinear system mapping, $x(t)$, $y(t)$ and $e(t)$ are the system input, output variables and the prediction error, respectively, n_x and n_y are the maximum input and output lags, respectively. And the observation noise $e(t)$ is an uncorrelated zero mean noise sequence providing that the function $f(\cdot)$ gives a sufficient description of the system. If the function $f(\cdot)$ is specified as a polynomial function, model (6.16) can then be represented by

$$y(t)=f\left(\varphi(t)\right)+e(t), \quad (6.17)$$

where $\varphi(t)=\left[y(t-1),\cdots,y(t-n_y),x(t-1),\cdots,x(t-n_x)\right]^T$ is the process regressor vector. The polynomial NARX model is a special case of the polynomial NARMAX model (Chen et al. 1990; Chen et al. 1989). The non-linear mapping $f(\cdot)$ of Eq. (6.17) can also be constructed using a class of local or global basis functions including radial basis functions (RBF), kernel functions, neural networks, multi-resolution wavelet such as B-splines and different types of polynomials such as the Chebyshev and Legendre types (Billings & Wei 2005a; Billings & Wei 2007; Billings et al. 2007; Chen et al. 1990; Chon et al. 2005; Harris et al. 2002; Li et al. 2011a; Liu 2001; Marinazzo et al. 2006; Niedzwiecki 1988; Pachori & Sircar 2008; Wei et al. 2009).

The polynomial bivariate model representation of NARX is represented by a compact matrix below

$$y(t)=\sum_{m=1}^{M}\alpha_m\Phi_m(t)+e_2(t), \tag{6.18}$$

where $\Phi_m(t)=\Phi_m(\varphi(t))$ are model terms generated from the regressor vector $\varphi(t)=\left[y(t-1),\cdots,y(t-n_y),x(t-1),\cdots,x(t-n_x)\right]^T$, α_m are unknown parameters and M is the total number of potential model terms. Note that the candidate model terms $\Phi_m(t)$ are of the form $x_1^{i_1}(t),\cdots,x_\lambda^{i_\lambda}(t)$, where λ refers to the nonlinear degree of the NARX model (6.18), $x_k^{i_k}(t)\in\{y(t-1),\cdots,y(t-n_y),x(t-1),\cdots,x(t-n_x)\}$ for $k=1,\cdots,\lambda$, $0\le i_k\le\lambda$ and $0\le i_1+\cdots+i_d\le\lambda$. The maximum lag of the polynomial model (6.17) is determined by n_y and n_x. The number of possible terms could be very large, and the number of polynomial terms (number of parameters) is $n_p=\dfrac{(n_y+n_x+\lambda)!}{(n_y+n_x)!\lambda!}$, for example, if $n_y=5$, $n_x=5$, and $\lambda=4$, $n_p=1001$. Particularly, if the non-linear degree λ of the NARX model (6-18) is reduced to 1, the NARX model simplifies to a linear ARX model described in Eq. (6.1) and Eq. (6.2).

The corresponding polynomial univariate NAR model can also be expressed by

$$y(t)=\sum_{m=1}^{M_0}\beta_m\Psi_m(t)+e_1(t), \tag{6.19}$$

where $\Psi_m(t)=\Psi_m(\varphi^*(t))$ are model terms generated from the regressor vector $\varphi^*(t)=\left[y(t-1),\cdots,y(t-n_y)\right]^T$, β_m are unknown parameters and M_0 is the total number of potential model terms.

The prediction error of the bivariate NARX model (6.18) (we assume $M>>n_y+n_x$) can be expressed by

$$\sum\nolimits_{y|y^-,x^-} = \frac{1}{M}\sum_{m=1}^{M}\left[y(t)-\alpha_m\Phi_m(t)\right]^2, \tag{6.20}$$

We can also consider the univariate NAR model (6.19) and obtain the corresponding prediction error

$$\sum\nolimits_{y|y^-} = \frac{1}{M_0}\sum_{m=1}^{M_0}\left[y(t)-\beta_m\Psi_m(t)\right]^2, \tag{6.21}$$

If the prediction of y improves by incorporating the past value of x, i.e., $\sum_{y|y^-,x^-}$ is smaller than $\sum_{y|y^-}$, then x is said to have a causal influence on y.

Modelling experience has shown that in most cases the initial full regression Eq. (6.18) might be highly redundant. Some of the regressors or model terms can be removed from the initial regression equation without any effect on the predictive capability of the model, and this elimination of the redundant regressors usually improves the model performance (Aguirre & Billings 1994). The ordinary least squares algorithm may fail to produce reliable parametric estimate results for such ill-posed problems. For most nonlinear dynamical system identification problems, only a relatively small number of model terms are commonly required in the regression model. Thus an efficient model term selection algorithm is highly desirable to detect and select the most significant regressors.

6.2.2.2 Model Structure Identification

The well-known orthogonal least squares (OLS) type of algorithms (Aguirre & Billings 1994; Aguirre & Billings 1995b; Billings & Wei 2007; Billings et al. 2007; Chen et al. 1989; Wei & Billings 2007; Wei et al. 2006; Wei et al. 2004b) have been proven to be very effective to solve multiple dynamical regression problems, where a great number of candidate model terms or regressors may be highly correlated and include in the regressor model. In the present study, the OLS algorithm discussed in (Billings & Wei 2007) is applied to deal with the regression model

(6.18). This involves a model refinement procedure including the selection of significant regressor or model terms.

6.2.2.3 Time-varying nonlinear model and parameter estimation

The time-varying (TV) VAR model fitting for bivariate NARX and univariate NAR model yields time-varying autoregressive parameters. Consequently, after the model refinement procedure, Eq. (6.18) and Eq. (6.19) are modified as follows:

TVNARX model: $$y(t)=\sum_{m=1}^{M^*}\alpha_m^*(t)\Phi_m^*(t)+e_2^*(t), \tag{6.22}$$

TVNAR model: $$y(t)=\sum_{m=1}^{M_0^*}\beta_m^*(t)\Psi_m^*(t)+e_1^*(t), \tag{6.23}$$

where M^*, M_0^* are the total number of the selected or significant regressors for the bivariate NARX and the univariate NAR model, respectively, ($M^*, M_0^* \le M$), α^* and β^* are time-varying parameters, Φ^* and Ψ^* are new model terms selected from the regression vector $\varphi(t)$ and $\varphi(t)^*$, e_2^* and e_1^* are the time-varying model prediction errors, respectively. An online recursive least squares (RLS) algorithm is then applied to estimate the time-varying model parameters. But other online methods, for example a Kalman filtering algorithm, can also be employed to estimate the unknown time-varying parameters. While the tracking ability of RLS is achieved by performing a forgetting operation on the information matrix, the tracking capability of Kalman filtering is obtained by adding a non-negative definite matrix to the covariance matrix. The main reason for employing RLS in the present study is because of its simple calculation and good convergence properties.

6.2.2.4 Time-varying nonlinear Granger causality measure

Similar to the definition of the linear Granger causality, let us begin with the bivariate case of causality $x \to y$. For model (6.18) and (6.19), the unbiased variance may be evaluated by the variance of prediction error described in Eq. (6.20) and (6.21). The time-varying estimation of VAR model (6.22) and (6.23) leads to time-varying prediction error. A general time-varying recursive variance computation is given in Eq. (6.13). If x causes y in the Granger causality sense, $\sum_{y|y^-,x^-}$ must then be smaller than $\sum_{y|y^-}$. Therefore, for the time-varying model (6.22) and (6.23), the calculation of time-varying nonlinear Granger causality can be evaluated by

$$NLGC_{x\to y}(t) = \ln \frac{\sum_{y|y^-}(t)}{\sum_{y|y^-,x^-}(t)}, \tag{6.24}$$

where $\sum_{y|y^-,x^-}(t)$ and $\sum_{y|y^-}(t)$ are the time-varying variance of the corresponding prediction error for the model (6.22) and (6.23), respectively. Exchanging the two time series, one may analogously study the time-varying nonlinear Granger causality influence of y on x. It is worth stressing that, within the definition of causality, for the time series data, directed flow of time plays a key role in making inference and depends on the direction. Note that Granger causality was initially formulated for linear models which may not be suitable for causality evaluation for nonlinear time series. Ancona et al. (2004) and Marinazzo et al. (2006) extended the Granger causality definition to nonlinear bivariate time series, and proposed that any prediction scheme providing a nonlinear extension of Granger causality should satisfy the following property: if a time series $\{x(t)\}_{t=1}^{N}$ is statistically independent of $\{y(t)\}_{t=1}^{N}$, then $\sum_{y|y^-,x^-} = \sum_{y|y^-}$; if $\{y(t)\}_{t=1}^{N}$ is statistically independent $\{x(t)\}_{t=1}^{N}$, then $\sum_{x|x^-,y^-} = \sum_{x|x^-}$, the property holds at least for $M \to \infty$. The polynomial NARX structure model is the class of nonlinear parametric models suitable to evaluate causality (Chen & Billings 1989). Ancona et al. (2004) introduced the nonlinear parametric model to evaluate the causality

based on a class of RBF models as a special case of a polynomial NARX structure model. It also should be noted that time-varying linear Granger causality is a special case of a time-varying NARX model to evaluate the Granger causality where the nonlinear degree d in the NARX model is equal to 1.

6.2.3 Choice of the model order

As to the issue of model order determination, this can be solved by using some model order determination criteria like Akaike information criterion (AIC), Bayesian information criterion (BIC) (Wei et al. 2010), Minimum description length (MDL) principle (Rissanen 1978), the Generalized cross-validation (GCV) criterion (Billings et al. 2007) or the visual fitting quality of the model (Brovelli et al. 2004).

6.2.4 Detection of nonlinearity in time series

In real-world applications, complex biomedical signal processes such as EEG time series probably comprise both linear and nonlinear components, together with stationary and nonstationary components. However, how to confirm whether the activity of the neural systems measured is regarded as linear/nonlinear or not is a challenging and controversial question. If the nature of the analysed signals was actually low-dimensional, the analytical results can be of great importance for theoretical neuroscience and clinical practice. However, confidence in results obtained from the detection of nonstationary nonlinearity including finite dimension, and positive Lyapunov exponents have also come into question, and alternative methods such as surrogate data technique for identifying possible nonlinear determinism in experimental data have been discussed and showed evidence of weak nonlinearity in nonstationary EEG time series (Palus 1996).

Although there are some different approaches to detecting the nonlinearity, the characterization of neural processes and brain signals from the mathematically tractable model is still an open issue. Some approaches for signal nonlinearity detection are usually set within a hypothesis-testing framework. Consequently, the rejection of the null hypothesis should be interpreted with due caution because the complex biomedical signals are subject to uncertainty and noise.

Therefore, in this Chapter, a delay vector variance (DVV) approach for the simultaneous characterization of nonstationary EEG signals in terms of nonlinearity is applied to detect the presence of nonlinearity (Mandic et al. 2008). For a given embedding dimension l and time series $\{x(t)\}_{t=1}^{N}$, the DVV procedure of detection for nonlinearity can be summarised as follows:

1) The mean μ_d, and standard deviation σ_d are computed over all pairwise distances between delay vectors, $\|x(i)-x(j)\|$ for $i \neq j$;
2) The set $\Omega_k(r_d)$ is generated subject to $\Omega_k(r_d)=\{x(i)\|x(k)-x(i)\| \leq r_d\}$, namely, the sets that consist of all delay vectors which lie closer to $x(k)$ than a certain distance r_d, taken from the interval $[\max\{0, \mu_d - n_d\sigma_d\}, \mu_d + n_d\sigma_d]$, where n_d is a parameter controlling the span over which to perform the DVV analysis;
3) For every set $\Omega_k(r_d)$, compute the variance of the corresponding targets $\sigma_k^2(r_d)$. The average over all sets $\Omega_k(r_d)$, normalized by the variance of the time series σ_x^2, yields the target variance $\sigma^{*2}(r_d)$ as following

$$\sigma^{*2}(r_d)=\frac{\frac{1}{N}\sum_{k=1}^{N}\sigma_k^2(r_d)}{\sigma_x^2}, \tag{6.25}$$

If the set $\Omega_k(r_d)$ contains at least 30 delay vectors, a valid variance measurement is considered, since having too few points for computing a sample variance yields unreliable estimates of the true variance. A sample of 30 data points for estimating a mean or variance is a general rule of thumb.

6.3 Simulation Example

In this section, we consider a simulation example that shows the ability of the time-varying Granger causality to react on changes in the directed influences between two signals. Consider the following time-varying ARX (2, 2) model:

$$\begin{aligned} x(t) &= a_{2,1}(t)x(t-1) + a_{2,2}(t)x(t-2) + c_{2,1}(t) \\ &\quad \times y(t-1) + c_{2,2}(t)y(t-2) + v_1(t), \\ y(t) &= b_{2,1}(t)y(t-1) + b_{2,2}(t)y(t-2) + d_{2,1}(t) \\ &\quad \times x(t-1) + d_{2,2}(t)x(t-2) + v_2(t), \end{aligned} \tag{6.27}$$

where

$$a_{2,1}(t) = \begin{cases} -0.6, & 1 \le t < 400, \\ 0.3, & 400 \le t \le 1000, \end{cases} \qquad b_{2,1}(t) = \begin{cases} 0.3, & 1 \le t < 400, \\ -0.6, & 400 \le t \le 1000, \end{cases}$$

$$a_{2,2}(t) = 0.1, \ 1 \le t \le 1000, \qquad b_{2,2}(t) = 0.1, \ 1 \le t \le 1000,$$

$$c_{2,1}(t) = \begin{cases} 0.2, & 1 \le t \le 300, \\ 0, & 300 < t \le 1000, \end{cases} \qquad d_{2,1}(t) = \begin{cases} 0, & 1 \le t < 700, \\ 0.2, & 700 \le t \le 1000, \end{cases} \tag{6.27}$$

$$c_{2,2}(t) = \begin{cases} 0.1, & 1 \le t \le 300, \\ 0, & 300 < t \le 1000, \end{cases} \qquad d_{2,2}(t) = \begin{cases} 0, & 1 \le t < 700, \\ 0.1, & 700 \le t \le 1000, \end{cases}$$

and

v_1, v_2 are Gaussian white noise processes with zero means and variances

$$\mathrm{var}(v_1) = \begin{cases} 0.9, & 1 \le t < 600, \\ 2.0, & 600 \le t \le 1000, \end{cases} \qquad \mathrm{var}(v_2) = \begin{cases} 2.0, & 1 \le t < 600 \\ 0.9, & 600 \le t \le 1000, \end{cases} \tag{6.28}$$

respectively. From the construction of the model, we can see, for the first 300

sample points, signal y causes signal x and, beginning with the sample point 700, signal x causes signal y. From sample point 301 up to sample point 699, no dependence exists between the two signals x and y. The results of time-varying Granger causalities are shown in Figure 6.1.

From Figure 6.1, the time-limited influence of $x \to y$ beginning with the sample index point 700 is detected by the positivity of $LGC_{x\to y}(t)$ (thick blue curve). The time-limited influence of $y \to x$ for the first 300 sample index points is identified by the positivity of $LGC_{y\to x}(t)$ (black curve). Obviously both Granger causalities are nearly zero within the time interval ($300 < t < 700$) without any dependence between the two signal components. Moreover, time-varying Granger causalities vary around the estimation of corresponding time-invariant Granger causality (for example, thin black and blue curves) within the stationary time intervals $0 < t \le 300$ and $700 \le t \le 1000$. In this simulation example, the time behaviour of time-varying Granger causality demonstrates the ability to react on changes in directed dependencies between two signals.

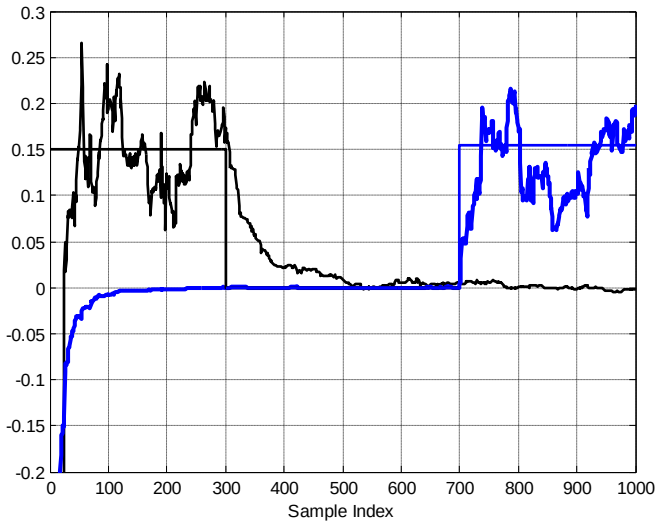


Figure 6.1 Time-varying Granger causalities $LGC_{x\to y}(t)$ (thick blue curve) and $LGC_{y\to x}(t)$ (black curve) from model (6.26) with time-varying parameters given in Eq. (6.27) are shown above. The thin blue and black step functions show that the correspondent estimated time-invariant Granger causalities $LGC_{x\to y}$ and $LGC_{y\to x}$.

6.4 Application to real EEG signal

A number of studies in the neuroscience literature have investigated the issue of causal effects in neural data (Brovelli et al. 2004; Ding et al. 2000; Gow et al. 2008; Marinazzo et al. 2006). In this example, we analyze a data set consisting of an epileptic sample of scalp EEGs recorded by Dr Sarrigiannis at the EEG Laboratory of Neurophysiology, Sheffield Teaching Hospitals NHS Foundation Trust, Royal Hallamshire Hospital.

6.4.1 Data acquisition

Scalp EEG signals are synchronous discharges from cerebral neurons detected by electrodes attached to the scalp. The EEG signals analyzed here were recorded by the same 32-channel amplifier system. A NeuroScan Medical System (NeuroSoft Inc., Sterling, VA) with the international 10-20 electrode coupling system (Rechtschaffen & Kales 1968) was used. An important issue in the EEG data acquisition is the problem of the reference electrode. There are several ways to define a reference electrode in scalp EEG recordings, as described in (Kaminski et al. 2001), not every type is suitable for the Granger causality analysis. Especially, the "common average" reference that involves all the channels as reference, and mixes signals from all of them. Generally, all operations where part of the signal from one channel appears in another channel will lead to spurious connections. In the present case study, the "bipolar montage" reference was used. Two examples are discussed in the real EEG application.

EEG data set 1: two bipolar montage channels "F3-C3" and "F4-C4" of EEG recorded from patient 1 with absence seizure epileptic discharge are investigated in this study, where the channel "F3-C3" represents the voltage difference between F3 and C3, and channel "F4-C4" means the voltage difference F4 and C4, respectively. The EEG signals between bipolar electrode channel "F3-C3" and channel "F4-C4" of 3500 data points pairs of one seizure, shown in Figure 6.2, which are for a sort of epileptic seizure activity of a patient, with a sampling rate of 500 Hz, recording during 7 seconds, were obtained for time-varying Granger causality analysis.

EEG data set 2: similarly, two bipolar montage channels "C3-T3" and "C4-T4" of EEG recorded from patient 2 with absence seizure epileptic discharge are also studied, where the channel "C3-T3" represents the voltage difference between C3 and T3, and channel "C4-T4" indicates the voltage difference C4 and T4, respectively. The EEG signals between bipolar electrode channel "C3-T3" and channel "C4-T4" of 5000 data points pairs of one seizure, shown in Figure 6-3, which are for a sort of epileptic seizure activity of a patient, with a sampling rate of

250 Hz, recording during 20 seconds, were obtained for time-varying Granger causality analysis. The reason for the selection of two group channels to analyse the causal influence is following advice from the clinician. Here the bipolar channel "F3-C3" indicates the new electrode channel F_3^*, the channel "F4-C4" means the new electrode channel F_4^*, and the bipolar channel "C3-T3" indicates the new electrode channel C_3^*, and the channel "C4-T4" means the new electrode channel C_4^*, respectively.

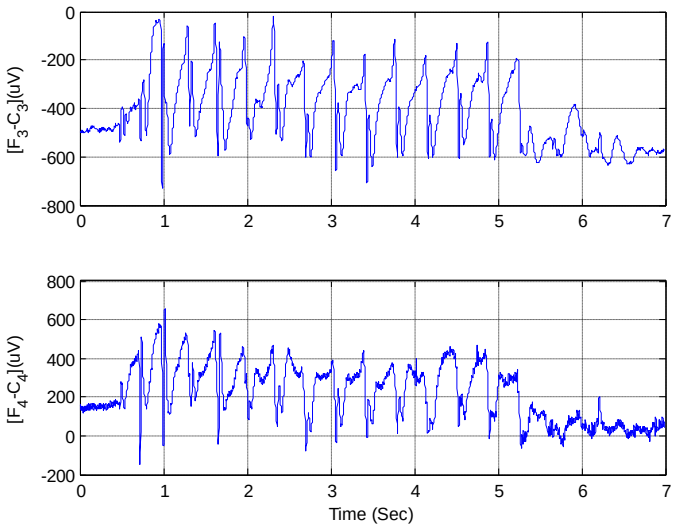


Figure 6.2 The EEG signals, for a sort of seizure activity of an epileptic patient 1, recorded during 7 seconds, with a sampling rate of 500 Hz, for both electrode channels F_3^* (above) and F_4^* (below).

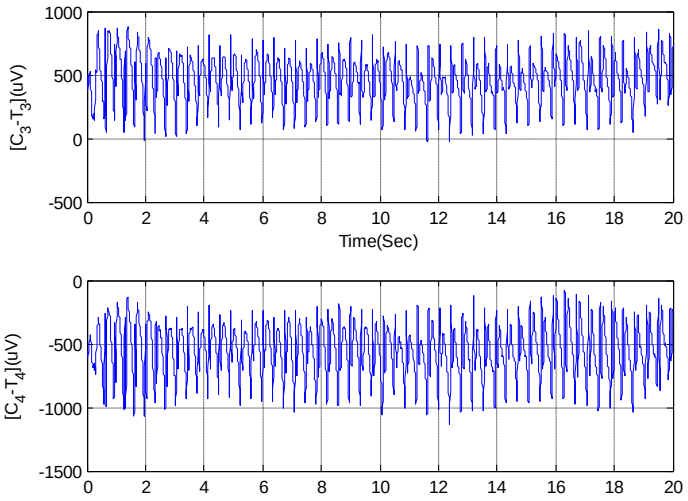
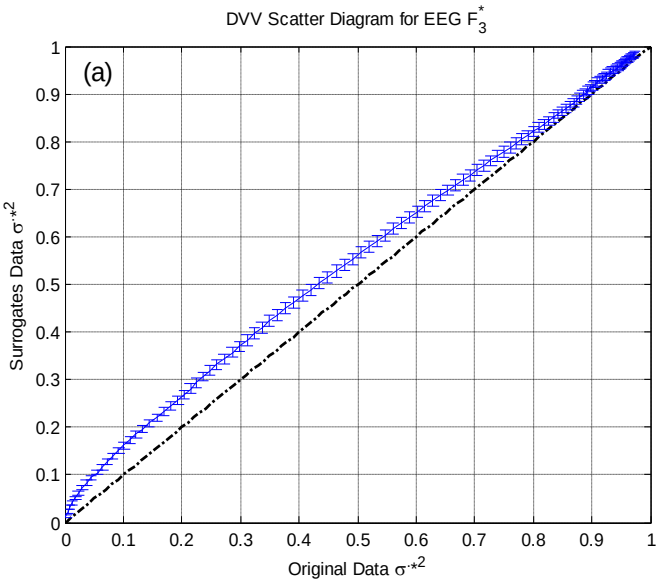


Figure 6.3 The EEG signals, for a sort of seizure activity of an epileptic patient 2, recorded during 20 seconds, with a sampling rate of 250 Hz, for both electrode channels C_3^* (above) and C_4^* (below).

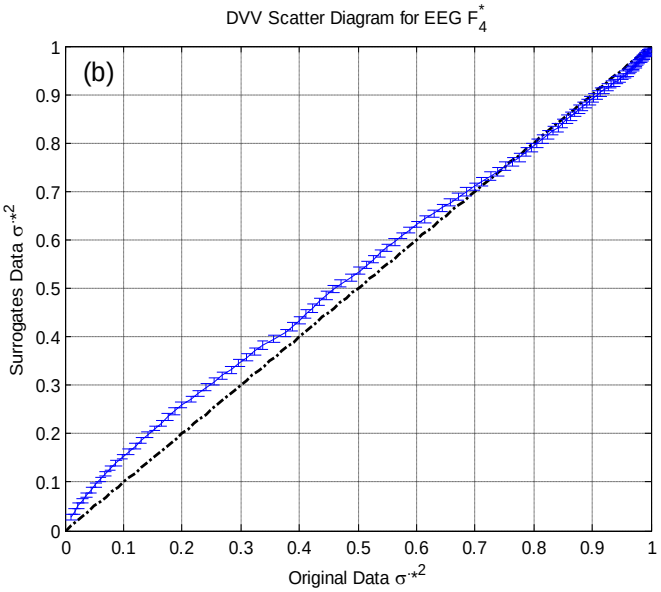
6.4.2 Nonlinear detection for EEG signals

The general time-varying linear model is still widely used in neuroscience (Li et al. 2011b; Pachori & Sircar 2008). Here our central aim is that the signal nonlinearity analysis should be undertaken prior to applying the proposed NARX modelling approach to real EEG signals for time-varying Granger causality. To achieve this, the DVV method discussed above is applied to the EEG signals acquired in section 6.4.1. The linear or nonlinear nature of the two bipolar EEG series is examined by performing DVV analyses on both the original and a number of surrogate time series, using the optimal embedding dimension of the original time series. Due to the standardisation of the distance axis, the DVV diagram can be combined in a scatter diagram, where the horizontal axis represents the original time series, and the vertical indicates the surrogate time series. If the surrogate time series is similar to the original series, the DVV scatter plot coincides with the bisector line, and the

original series is regarded as linear approximately. Conversely, the deviation from the bisector line is an indication of nonlinearity. In Figure 6.4, the DVV scatter diagrams for EEG signals F_3^* and F_4^* are nonlinear.



(a)



(b)

Figure 6.4 DVV scatter plot for the EEG time series given in Fig 6.2.

(a) The EEG Channel F_3^* (above),

(b) The EEG Channel F_4^* (below).

6.4.3 Time-varying nonlinear Granger causality for EEG signals

Generally, the most significant property of the Granger causality is its positivity, when a signal X causes a second signal Y. From the discussion in section 6.2.2.3, the time-varying NARX modelling approach under definition (6.22) and (6.23) is applied to real EEG signals to study the causal relationship for EEG signals F_3^* and F_4^*, C_3^* and C_4^*, respectively.

The Granger causality analysis for EEG data set 1. The NARX model with a non-linear degree $d=2$, maximum lags $n_y=5$ and $n_x=5$, and the total number of potential 66 regressor terms, for different model orders were estimated using the OLS algorithm (Billings & Wei 2007), and both the AIC and BIC criteria suggested that the model size can be chosen to be 7 from the total number of 66 regressor terms, with the bivariate case $F_3^* \to F_4^*$ ($x \to y$), Hence time-varying NARX model and univariate NAR model can be represented by, respectively

TVNARX: $$y(t)=\sum_{i=1}^{3}\theta_{1,i}(t)y(t-i)+\theta_{1,1}(t)x(t-1)+\theta_{2,1}(t)x(t-1)+\theta_{2,5}(t)x(t-5)+\theta_{5,5}(t)x^2(t-5), \quad (6.29)$$

TVNAR: $$y(t)=\sum_{i=1}^{3}\theta_i^*(t)y(t-i)+\theta_{1,1}^*(t)x(t-1), \quad (6.30)$$

Due to Eq. (6.29), (6.30) and (6.13), the transient estimations of time-varying nonlinear Granger causality $NLGC_{F_3^* \to F_4^*}(t)$ can be obtained. Similarly, exchanging the two EEG signals, for bivariate $F_4^* \to F_3^*$, the time-varying Granger causality $NLGC_{F_4^* \to F_3^*}(t)$ can also be obtained for both cases for $c=0.01$. The time-varying bivariate and univariate model for case $F_4^* \to F_3^*$ ($y \to x$) can be

shown as follows.

$$x(t)=\theta_0+\sum_{i=1}^{5}\theta_{1,i}x(t-i)+\theta_{2,1}y(i-1)+\theta_{2,5}x^2(i-5), \tag{6.31}$$

$$x(t)=\theta_0'+\sum_{i=1}^{5}\theta_{1,i}'x(t-i)+\theta_{2,5}'x^2(i-5), \tag{6.32}$$

Time-varying Granger causality in both directions was calculated for electrode pairs between F_3^* and F_4^*. A directed influence for a determined time interval is stated. Figure 6.5 illustrates a typical result for the electrode pair F_3^*/F_4^*. In Figure 6.5, we depict the directed interactions $NLGC_{F_3^*\to F_4^*}(t)$ (blue curve, measuring the influence of F_3^* on F_4^*) and $NLGC_{F_4^*\to F_3^*}(t)$ (black curve, measuring the influence of F_4^* on F_3^*), as a function of sample time, for a sort of epileptic seizure activity of a patient. From Figure 6.5, two directed Granger causalities can be clearly observed. 1) The Granger causality $NLGC_{F_3^*\to F_4^*}(t)$ is significantly larger in the time interval 0.5-2s than $NLGC_{F_4^*\to F_3^*}(t)$. Thus, a superior influence from F_3^* to F_4^* is present within this time interval, especially, compared to the original EEG time series given in Figure 6.2, there is a very strong Granger causality influence observed at the time point of 1 second; 2) during the period from 3.5 to 3.8s, the Granger causality influence from F_4^* to F_3^* is dominant. It is also worth noting that these causal relationships are not evidenced in terms of cross correlation which is defined as

$$c_1(\tau)=\frac{\sum_t[x(t-\tau)-\bar{x}][y(t)-\bar{y}]}{\sqrt{\sum_t[x(t)-\bar{x}]^2}\sqrt{\sum_t[y(t)-\bar{y}]^2}}, \tag{6.33}$$

and

$$c_2(\tau)=\frac{\sum_t[x(t+\tau)-\bar{x}][y(t)-\bar{y}]}{\sqrt{\sum_t[x(t)-\bar{x}]^2}\sqrt{\sum_t[y(t)-\bar{y}]^2}}, \tag{6.34}$$

$c_1(\tau)$ and $c_2(\tau)$ for some specific epileptic patient data are depicted in Figure 6.6,

which gives no interesting patterns that are possessed by the proposed time-varying NARX model.

The Granger causality analysis for EEG data set 2 is similar to the discussion of EEG data set 1. Hence the time-varying NARX model and the univariate NAR model with the bivariate case $C_3^* \to C_4^*$ ($x \to y$) can be represented by, respectively

TVNARX: $$y(t) = \sum_{i=1}^{3} \theta_{1,i}\, y(t-i) + \theta_{5,1}\, y(t-5)$$

$$\times x(t-2) + \theta_{3,1} y(t-1) x(t-1), \tag{6.35}$$

TVNAR: $$y(t) = \sum_{i=1}^{3} \theta'_{1,i} y(t-i), \tag{6.36}$$

Due to Eq. (6.35), (6.36) and (6.13), the transient estimations of time-varying nonlinear Granger causality $NLGC_{C_3^* \to C_4^*}(t)$ can be obtained. Similarly, exchanging the two EEG signals, for bivariate $C_4^* \to C_3^*$, the time-varying Granger causality $NLGC_{C_4^* \to C_3^*}(t)$ can be also obtained for both cases for $c = 0.005$. Also, the time-varying bivariate and univariate model for case $C_4^* \to C_3^*$ ($y \to x$) can be represented by

$$x(t) = \sum_{i=1}^{3} \theta_{1,i} x(t-i) + \theta_{2,5} y^2(t-5) + \theta_{3,1} y^2(t-1) + \theta_{3,2} y^2(t-2), \tag{6.37}$$

$$x(t) = \sum_{i=1}^{3} \theta'_{1,i} x(t-i), \tag{6.38}$$

Time-varying Granger causality in both directions was calculated for electrode pairs between C_3^* and C_4^*. A directed influence for a determined time interval is stated. Figure 6.7 illustrates a typical result for the electrode pair C_3^* / C_4^*. In Figure 6.7, we depict the directed interactions $NLGC_{C_3^* \to C_4^*}(t)$ (blue curve,

measuring the influence of C_3^* on C_4^*) and $NLGC_{C_4^*\to C_3^*}(t)$ (black curve, measuring the influence of C_4^* on C_3^*), as a function of sample time, for a sort of epileptic seizure activity of a patient. From Figure 6.7, two directed Granger causality can be clearly observed. 1) For the chosen electrode pair C_3^*/C_4^*, the interaction is directed from left central area (C_3^*) to right central area (C_4^*) during the whole time interval, and the Granger causality $NLGC_{C_3^*\to C_4^*}(t)$ is significantly larger than $NLGC_{C_4^*\to C_3^*}(t)$ in the whole of the time interval. Thus, a superior influence from C_3^* to C_4^* is present within this time interval; 2) during the period from 10 to 12s, the Granger causality influence between C_4^* and C_3^* is weaker, especially, about the time point 20s, the interaction is very small. It is worth noting that these causal relationships for the electrode pair C_3^*/C_4^* are not evident in terms of cross correlation that the cross correlation can only be applied to detect the relationship between EEG signals in linear model case. However, the identified model for the EEG signals is nonlinear model. The cross correlation functions $c^*_1(\tau)$ and $c^*_2(\tau)$ for both electrode, depicted in Figure 6-8 for epileptic patient data, do not show interesting patterns. These properties, possessed by the proposed time-varying NARX model, cannot be obtained using any time-invariant parametric modelling framework for linear and nonlinear Granger causality of time series. It should also be stressed that, in the case of an optimal fit to the true autoregressive parameters of univariate and bivariate models, Granger causality is a non-negative value. Compared to the Granger causality analysis of the occurrence of negative values mainly due to not having optimal models in (Hesse et al. 2003), our proposed approach in this study is novel and effective. The results can help clinicians interpret EEG signals.

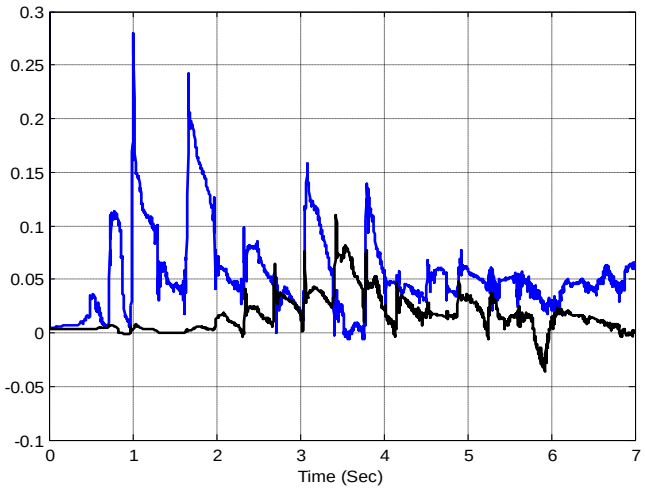


Figure 6.5 The time-varying non-linear Granger causality $NLGC_{F_3^* \to F_4^*}(t)$ (blue curve) $NLGC_{F_4^* \to F_3^*}(t)$ (black curve) and are plotted versus the time courses for EEG signals shown in Figure 6.2.

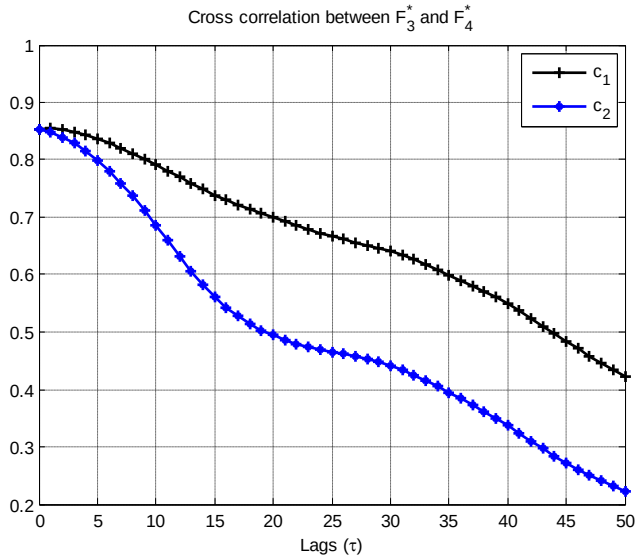


Figure 6.6 The cross correlations $c_1(\tau)$ (black curve) and $c_2(\tau)$ (blue curve) are shown versus τ between EEG signals F_3^* and F_4^* given in Figure 6.2.

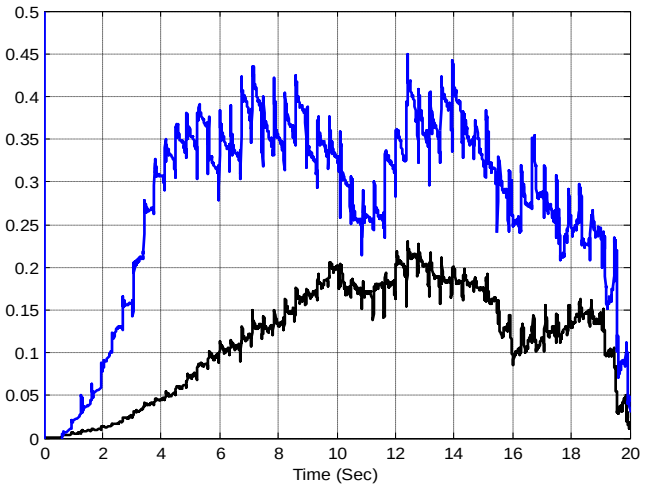


Figure 6.7 The time-varying non-linear Granger causality $NLGC_{C_3^* \to C_4^*}(t)$ (blue curve) $NLGC_{C_4^* \to C_3^*}(t)$ (black curve) and are plotted versus the time courses for EEG signals shown in Figure 6.3.

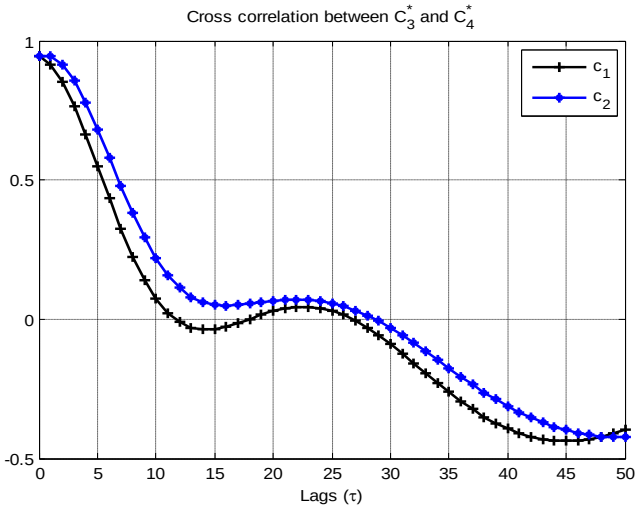


Figure 6.8 The cross correlations $c_1'(\tau)$ (black curve) and $c_2'(\tau)$ (blue curve) are shown versus τ between EEG signals C_3^* and C_4^* given in Figure 6.3.

6.5 Conclusions

In this Chapter we have introduced the polynomial mathematical formalism and studied the problem of how to evaluate time-varying linear and nonlinear Granger causal relations in neural systems. Demonstrations of the technique have been carried out on both simulated data, where the patterns of interactions are known, and on real EEG signals. This Chapter mainly illuminates four essential aspects as follows. First, the proposed method for time-varying linear estimation of Granger causality permits the detection of temporal causal interactions. Second, we have generalized the nonlinear parametric approach to Granger causality: the proposed model can approximate any functions such as RBF, neural networks, multi-resolution wavelet and different types of polynomials including the Chebyshev and Legendre types. Third, prior to applying the method to the actual EEG data to analyse the time-varying Granger causality, a delay vector variance (DVV) approach is applied to detect the presence of nonlinearity in the EEG signals and show that the nonstationary EEG recording convey nonlinearity. Finally, temporally directed interactions were detected successfully for electrophysiological data of two epileptic patients on the basis of transient Granger causality.

This Chapter demonstrates the possibility of the detection and description of transient directed interactions for the bivariate case. In fact, the presented approach can be extended to the multivariate case and be suitable for the study of causal interactions between electrophysiological data of different sites on the scalp, occurring during the cognitive processes.

Chapter 7

Conclusions and Future Work

This chapter summaries the main work and key contributions of the thesis, and then provides some suggestions for future study.

7.1 Contributions

The main purpose of this thesis is that modelling and adaptive tracking of nonstationarities for both linear and nonlinear systems has been investigated on the basis of the NARMAX model and wavelet basis expansion in both time and frequency domains. The work of this thesis focuses on the investigation of methods and algorithms for modelling and tracking nonstationary processes of both linear and nonlinear systems in both the time and frequency domains. The main contributions are summarized below:

- **Linear modelling and adaptive tracking:** In Chapter 3 we introduced a new parametric modelling and identification method for linear time-varying systems using a block least mean square (LMS) approach where the time-varying parameters are approximated using multi-wavelet basis functions. This approach can be applied to track rapidly or even sharply varying processes and is developed by combining wavelet approximation theory with a block LMS algorithm. Numerical examples are provided to show the effectiveness of the proposed method for dealing with severely nonstationary processes. Application of the proposed approach to a mechanical system indicates better tracking capability of the multi-wavelet basis function algorithm compared with the normalized least squares or recursive least

squares routines.

- **Time-frequency feature extraction:** In Chapter 4 a novel modelling scheme that can be used to estimate and track time-varying properties of nonstationary signals was investigated. This scheme is based on a class of time-varying autoregressive with exogenous input (TVARX) models where the associated time-varying parameters are represented by multi-wavelet basis functions. The orthogonal least square (OLS) algorithm is then applied to refine the model parameter estimates of the TVARX model. The main features of the multi-wavelet approach is that it enables smooth trends to be tracked but also to capture sharp changes in the time-varying process parameters. Simulation studies and applications to real EEG data show that the proposed algorithm can provide important transient information on the inherent dynamics of nonstationary processes.

- **CMSS structure algorithm**: In Chapter 5 the identification of nonlinear time-varying systems using linear-in-the-parameter models was investigated. A new efficient common model structure selection (CMSS) algorithm is proposed to select a common model structure. The main idea and key procedure is first to generate $K+1$ data sets (the first K data sets are used for training, and the $(K+1)$th one is used for testing) using an online sliding window method; then detect significant model terms to form a common model structure which fits over all the K training data sets using the new proposed CMSS approach, and finally, estimate and refine the time-varying parameters for the identified common-structured model using sliding-window recursive least squares (SWRLS) approach. The new method can effectively detect and adaptively track the transient variation of nonstationary signals. Two examples are presented to illustrate the effectiveness of the new approach including an application to an EEG data set.

- **Time-varying linear and nonlinear Granger causality:** In Chapter 6, mathematical measures such as coherence, mutual information or correlation are usually applied to evaluate and describe the interactions between groups of neurons to investigate neural connections. However, these methods cannot distinguish directions of flow between two cortical sites or causality. Being able to assess the directionality of neuronal interactions is thus a highly desired capability for understanding the cooperative nature of neural computation. Granger causality, which is a fundamental tool for the description of causal interaction of two signals, is a key technique to furnish this capability. The main topic of this Chapter is that a novel linear and nonlinear time-varying parametric modelling and identification approach using data-driven methods was proposed for the adaptive estimation of nonstationary EEG Granger causality processes to detect the transient dynamical causal directional interactions between EEG signals within time intervals in the time domain. The time-varying model proposed allows identification of the direction of information flow between brain areas, extending the Granger causality concept to transient time-varying processes. A numerical example, where the exact answers of causal influences are known, demonstrates a good performance of the time-varying Granger causality for detecting transient dynamical causal relations over the time course. The application of this novel approach is then applied to analyse EEG signal local field potentials to track and detect the causal influences between EEG signals. One advantage of the proposed model, compared with traditional Granger causality assuming the stationarity of the signals, is that our results can be more interpretable and yield new insights into the transient directed dynamical Granger causality interactions.

7.2 Future Work

Data-based identification and modelling of nonstationary signals are very complex

and difficult. Although there are some powerful methods such as the adaptive algorithms and the OLS method, some important issues still need to be investigated further, for example,

- As mentioned in Chapter 4, the time varying AR and ARX identification and modelling method based on multi-wavelet basis function expansion investigated outperforms the time invariant modelling approach. Furthermore these methods should be extended to the time-varying NARMAX model and to capture and extract the time-frequency properties in the time-frequency domain. It should be noted that the existing generalised frequency response functions (GFRFs) can only analyse the time-invariant models estimated in the frequency domain.

- These methods can be extended to a time-varying nonlinear autoregressive (TVNAR) model, but mapping to the frequency domain may be difficult because the time-varying nonstationary signal investigated can be treated as not Gaussian, and higher-order statistical information to deal with such things as degrees of nonlinearity and deviations from normality, should be considered. The higher-order spectrum (HOS), called the polyspectrum, which is defined in terms of the higher-order statistics of a signal, can provide additional information and, in some applications, has been proved to be a very effective tool, especially for ‘non’ processes and systems: non-Gaussian, non-linear, non-stationary, non-minimum phase, non-causal and non-additive ones.

- In Chapter 5, the common model structure selection approach has been investigated and designed to identify a robust time-varying common-structured (TVCS) model as a solution to time-varying nonlinear systems identification problems using an online sliding-window approach. Furthermore, the TVCS mode can also be estimated on the basis of multi-wavelet basis function expansion to improve the tracking ability of the nonstationary systems.

- In Chapter 6, the time-varying linear and nonlinear Granger causality in the time domain was presented. The directed detection of Granger causality on the basis of basis function expansion should also be studied to improve the detection capability of Granger causality. Especially, the issue of occurrence of negative values may be solved due to the non-consistency of the model parameters estimation and fluctuation from the RLS algorithm that the negative values of estimated Granger causality are possible. Another future work is that the measure causality in Chapter 6 is not only considered between two components, but we are able to estimate autoregressive models for an arbitrary number of time series, the direct or indirect influences can also be identified from other components.

- Regarding nonstationarities and nonlinearity detection, electrophysiological data including EEG data are generally regarded as non-stationary because their statistical characteristics change over the time course, depending on the mental states which are active at any time instant. For the issue of nonlinearities of the time series, some authors have investigated using different detection criteria such as largest Lyapunov exponent, correlation integral, delay vector variance (DVV) and time-varying surrogate data to detect the presence of nonlinear dynamics. In our future work, we can use the time-varying NARMAX methodology to test the presence of nonlinear dynamics in nonstationary time series, and hopefully obtain more interesting results compared to the existing detection approaches.

Papers Arising from This Thesis

Li, Y., Wei, H. L. and Billings, S. A. (2011): Identification of time-varying systems using multi-wavelet basis functions. *IEEE Transactions on Control Systems Technology* 19, 656-663.

Li, Y., Wei, H. L., Billings, S. A. and Sarrigiannis, P. (2011): Time- varying model identification for time-frequency feature extraction from EEG data. *Journal of Neuroscience Methods* 196, 151-158.

Li, Y., Wei, H. L., Billings, S. A. and Sarrigiannis, P. Identification of nonlinear time-varying systems using an online sliding-window and common model structure selection (CMSS) approach with applications to EEG, *Department of Automatic Control and Systems Engineering, The University of Sheffield, Research Report* (No. 1019).

Li, Y., Wei, H. L., Billings, S. A., Balikhin, M. A. and Walker, S. A. (2009): A Time varying model for Disturbance Storm Time (Dst) Index analysis. *The 9th UK CARE Annual General Meeting, Proceeding, Manchester*, September 2009.

Li, Y., Wei, H. L. and Billings, S. A. (2011): A Time-varying model for EEG Granger Causality analysis. *The 4th Biennial Neuroscience Conference, Germany*, May 2011.

Bibliography

Adeli H, Zhou Z, Dadmehr N. (2003). Analysis of EEG records in an epileptic patient using wavelet transform. *Journal of neuroscience methods* 123: 69-87

Aguirre LA, Billings SA. (1994). Validating identified nonlinear models with chaotic dynamics. *Int J Bifurcat Chaos* 4: 109-125

Aguirre LA, Billings SA. (1995a). Dynamical effects of overparametrization in nonlinear models. *Physica D* 80: 26-40

Aguirre LA, Billings SA. (1995b). Retrieving dynamical invariants from chaotic data using NARMAX models. *Int J Bifurcat Chaos* 5: 449-474

Ahuja N, Lertrattanapanich S, Bose NK. (2005). Properties determining choice of mother wavelet. *IEE Proceedings-Vision Image and Signal Processing* 152: 659-664

Akaike H. (1974). New Look at Statistical-Model Identification. *IEEE Transactions on Automatic Control* AC19: 716-723

Ancona N, Marinazzo D, Stramaglia S. (2004). Radial basis function approach to nonlinear Granger causality of time series. *Physical Review E* 70

Andrzejak RG, Lehnertz K, Mormann F, Rieke C, David P, Elger CE. (2001). Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E* 64

Ansari-Asl K, Bellanger JJ, Bartolomei F, Wendling F, Senhadji L. (2005). Time-frequency characterization of interdependencies in nonstationary signals: Application to epileptic EEG. *IEEE Transactions on Biomedical Engineering* 52: 1218-1226

Belge M, Miller EL. (2000). A sliding window RLS-like adaptive algorithm for filtering alpha-stable noise. *IEEE Signal Processing Letters* 7: 86-89

Billings SA, Chen S, Korenberg MJ. (1989). Identification of MIMO non-linear systems using a forward-regression orthogonal estimator. *International Journal of Control* 49: 2157-2189

Billings SA, Coca D. (1999). Discrete wavelet models for identification and qualitative

analysis of chaotic systems. *Int J Bifurcat Chaos* 9: 1263-1284

Billings SA, Jones JCP. (1990). Mapping nonlinear integrodifferential equations into the frequency - domain. *International Journal of Control* 52: 863-879

Billings SA, Voon WSF. (1986). Correlation Based Model Validity Tests for nonlinear Models. *International Journal of Control* 44: 235-244

Billings SA, Wei HL. (2005a). A new class of wavelet networks for nonlinear system identification. *IEEE Transactions on Neural Networks* 16: 862-874

Billings SA, Wei HL. (2005b). The wavelet-NARMAX representation: A hybrid model structure combining polynomial models with multiresolution wavelet decompositions. *International Journal of Systems Science* 36: 137-152

Billings SA, Wei HL. (2007). Sparse model identification using a forward orthogonal regression algorithm aided by mutual information. *IEEE Transactions on Neural Networks* 18: 306-310

Billings SA, Wei HL, Balikhin MA. (2007). Generalized multiscale radial basis function networks. *Neural Networks* 20: 1081-1094

Billings SA, Zhu QM. (1994). Nonlinear Model Validation Using Correlation Tests. *International Journal of Control* 60: 1107-1120

Blinowska KJ, Kus R, Kaminski M. (2004). Granger causality and information flow in multivariate processes. *Physical Review E* 70

Bouzeghoub MC, Ellacott SW, Easdown A, Brown M. (2000). On the identification of non-stationary linear processes. *International Journal of Systems Science* 31: 273-286

Brovelli A, Ding MZ, Ledberg A, Chen YH, Nakamura R, Bressler SL. (2004). Beta oscillations in a large-scale sensorimotor cortical network: Directional influences revealed by Granger causality. *Proceedings of the National Academy of Sciences of the United States of America* 101: 9849-9854

Burke DR, Kelly SR, de Chazal P, Reilly RB, Finucane C. (2005). A parametric feature extraction and classification strategy for brain-computer interfacing. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 13: 12-17

Cakrak F, Loughlin PJ. (2001). Multiple window time-varying spectral analysis. *IEEE Transactions on Signal Processing* 49: 448-453

Chang CM, Liu TS. (2006). A wavelet network control method for disk drives. *IEEE Transactions on Control Systems Technology* 14: 63-68

Chen S, Billings SA. (1989). Representations of non-linear systems - the NARMAX model. *International Journal of Control* 49: 1013-1032

Chen S, Billings SA, Cowan CFN, Grant PM. (1990). Practical identification of NARMAX models using radial basis functions. *International Journal of Control* 52: 1327-1350

Chen S, Billings SA, Luo W. (1989). Orthogonal Least-Squares Methods and Their Application to Non-linear System-Identification. *International Journal of Control* 50: 1873-1896

Chen S, Wang XX, Harris CJ. (2008). NARX-based nonlinear system identification using orthogonal least squares basis hunting. *IEEE Transactions on Control Systems Technology* 16: 78-84

Chen SL, Lai HC, Ho KC. (2006). Identification of linear time varying systems by Haar wavelet. *International Journal of Systems Science* 37: 619-628

Chen W, Chowdhury FN. (2007). Simultaneous identification of time-varying parameters and estimation of system states using iterative learning observers. *International Journal of Systems Science* 38: 39-45

Choi BY, Bien Z. (1989). Sliding-windowed weighted recursive least-squares method for parameter-estimation. *Electronics Letters* 25: 1381-1382

Chon KH, Zhao H, Zou R, Ju K. (2005). Multiple time-varying dynamic analysis using multiple sets of basis functions. *IEEE Transactions on Biomedical Engineering* 52: 956-960

Chowdhury FN. (2000). Input-output Modelling of nonlinear systems with time-varying linear models. *IEEE Transactions on Automatic Control* 45: 1355-1358

Chowdhury FN, Aravena JL. (1998). A modular methodology for fast fault detection and classification in power systems. *IEEE Transactions on Control Systems Technology* 6: 623-634

Chowdhury FN, Jiang B, Belcastro CM. (2006). Reduction of false alarms in fault detection problems. *International Journal of Innovative Computing Information and Control* 2: 481-490

Chui CK. (1992). *An Introduction to Wavelets*. Boston: Academic Press.

Clark GA, Mitra SK, Parker SR. (1981). Block implementation of adaptive digital - filters. *IEEE Transactions on Circuits and Systems* 28: 584-592

Cohen A. (1986). *Biomedical Signal Processing*. Boca Raton, Fla, CRC Press.

Cohen A. (2000). *Biomedical signals: origin and dynamic characteristics; frequency-domain analysis*. The Biomedical Engineering Handbook, second ed., CRC Press LLC, Boca Raton, FL,.

Cohen L. (1995). *Time-frequency analysis: theory and applications.* Prentice-Hall, Inc. Upper Saddle River, NJ, USA.

Ding MZ, Bressler SL, Yang WM, Liang HL. (2000). Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive Modelling: data preprocessing, model validation, and variability assessment. *Biological Cybernetics* 83: 35-45

Fallah S, Ramachandran V. (1992). A New efficient algorithm for adaptive lattice filters. *Comput. Electr. Eng.* 18: 217-225

Farmer JD, Sidorowich JJ. (1987). Predicting chaotic time-series. *Physical Review Letters* 59: 845-848

Fase L, Zhao H, Chon KH, Nollo G. (2009). Time-Varying Surrogate Data to Assess Nonlinearity in Nonstationary Time Series: Application to Heart Rate Variability. *IEEE Transactions on Biomedical Engineering*

Forte JD, Bui BV, Vingrys AJ. (2008). Wavelet analysis reveals dynamics of rat oscillatory potentials. *Journal of neuroscience methods* 169: 191-200

Fortescue TR, Kershenbaum LS, Ydstie BE. (1981). Implementation of Self-Tuning Regulators with Variable Forgetting Factors. *Automatica* 17: 831-835

Ghanem R, Romeo F. (2000). A wavelet-based approach for the identification of linear time-varying dynamical systems. *J Sound Vib* 234: 555-576

Goodwin GC, Sin KS. (1984). *Adaptive Filtering, Prediction, and Control*. Englewood Cliffs, NJ: Prentice-Hall.

Gourevitch B, Le Bouquin-Jeannes R, Faucon G. (2006). Linear and nonlinear causality between signals: methods, examples and neurophysiological applications. *Biological*

Cybernetics 95: 349-369

Gow DW, Segawa JA, Ahlfors SP, Lin FH. (2008). Lexical influences on speech perception: A Granger causality analysis of MEG and EEG source estimates. *NeuroImage* 43: 614-623

Granger CWJ. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37: 424-438

Harris C, Hong X, Gan Q. (2002). *Adaptive modelling, estimation, and fusion from data: a neurofuzzy approach*. Springer Verlag.

Hayes MH. (1996). *Statistical digital signal processing and modelling.*: New York [etc.] : John Wiley & Sons.

Haykin S. (2003). *Adaptive filter theory (ISE)*. Prentice-Hall, Englewood-Cliffs, NJ.

Hesse W, Moller E, Arnold M, Schack B. (2003). The use of time-variant EEG Granger causality for inspecting directed interdependencies of neural assemblies. *Journal of neuroscience methods* 124: 27-44

Huang NE, Shen Z, Long SR, Wu MLC, Shih HH, et al. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences* 454: 903-995

Huang NE, Wu Z, Long SR. (2008). Hilbert-Huang transform. *Scholarpedia* 3: 2544

Hwang DH, Schmitt WA, Stephanopoulos G. (2002). Determination of minimum sample size and discriminatory expression patterns in microarray data. *Bioinformatics* 18: 1184-1193

Jiang J, Cook R. (1992). Fast Parameter Tracking RLS Algorithm with High Noise-Immunity. *Electronics Letters* 28: 2043-2045

Johansson R. (1993). *System Modelling and Identification*. Englewood Cliffs, NJ: Prentice-Hall.

Jones JCP, Billings SA. (1989). Recursive algorithm for computing the frequency - response of a class of non-linear difference equation models. *International Journal of Control* 50: 1925-1940

Kaipio JP, Karjalainen PA. (1997). Estimation of event-related synchronization changes by

a new TVAR method. *IEEE Transactions on Biomedical Engineering* 44: 649-656

Kaminski M, Ding MZ, Truccolo WA, Bressler SL. (2001). Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biological Cybernetics* 85: 145-157

Kay SM. (1988). *Modern Spetral Estimation.* Prentice Hall, Englewood Cliffs, NJ.

Korenberg M, Billings SA, Liu YP, McIlroy PJ. (1988). Orghogonal Parameter-Estimation Algorithm for Non-linear Stochastic-Systems. *International Journal of Control* 48: 193-210

Leontaritis IJ, Billings SA. (1985). Input output parametric models for non-linear systems .1. Deterministic non-linear systems. *International Journal of Control* 41: 303-328

Leontaritis IJ, Billings SA. (1987). Experimental - design and identifiability for nonlinear - systems. *International Journal of Systems Science* 18: 189-202

Leung SH, So CF. (2005). Gradient-based variable forgetting factor RLS algorithm in time-varying environments. *IEEE Transactions on Signal Processing* 53: 3141-3150

Li Y, Wei HL, Billings SA. (2011a). Identification of Time-Varying Systems Using Multi-Wavelet Basis Functions. *IEEE Transactions on Control Systems Technology* 19: 656-663

Li Y, Wei HL, Billings SA, Sarrigiannis PG. (2011b). Time-varying model identification for time-frequency feature extraction from EEG data. *Journal of neuroscience methods* 196: 151-158

Liu GP. (2001). *Nonlinear identification and control: a nueral network approach*. Springer Verlag.

Ljung L. (1999). *System Identification: Theory for the User*. 2nd ed. Upper Saddle River, NJ: Prentice-Hall.

Ljung L. (2001). *Black-box models from input-output measurements.* Presented at Imtc/2001: Proceedings of the 18th IEEE Instrumentation and Measurement Technology Conference, Vols 1-3,

Ljung L, Gunnarsson S. (1990). Adaptation and tracking in system-identification - a survey. *Automatica* 26: 7-21

Mallat S. (1999). *A Wavelet Tour of Signal Processing*. Academic Press, London.

Mallat SG. (1989). A theory for multiresolution signal decomposition - the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11: 674-693

Mandic DP, Chen M, Gautama T, Van Hulle MM, Constantinides A. (2008). On the characterization of the deterministic/stochastic and linear/nonlinear nature of time series. *Proceedings of the Royal Society a-Mathematical Physical and Engineering Sciences* 464: 1141-1160

Manolakis DG, Ingle VK, Kogon SM, ebrary I. (2005). *Statistical and adaptive signal processing: spectral estimation, signal modelling, adaptive filtering, and array processing*. Artech House.

Marinazzo D, Pellicoro M, Stramaglia S. (2006). Nonlinear parametric model for Granger causality of time series. *Physical Review E* 73

Morbidi F, Garulli A, Prattichizzo D, Rizzo C, Rossi S. (2008). Application of Kalman Filter to Remove TMS-Induced Artifacts from EEG Recordings. *IEEE Transactions on Control Systems Technology* 16: 1360-1366

Mukhopadhyay S, Sircar P. (1997). Parametric modelling of non-stationary signals: A unified approach. *Signal Processing* 60: 135-152

Nagumo JI, Noda A. (1967). A learning method for system identification. *IEEE Transactions on Automatic Control* AC12: 282-&

Nells O. (2001). *Nonlinear System Identification: From Classical Approach to Neural Networks and Fuzzy Models*. Springer-Verlag, Berlin: Heidelberg.

Ng SC, Raveendran P. (2009). Enhanced mu Rhythm Extraction Using Blind Source Separation and Wavelet Transform. *IEEE Transactions on Biomedical Engineering* 56: 2024-2034

Niedzwiecki M. (1988). Functional Series Modelling Approach to Identification of Nonstationary Stochastic-Systems. *IEEE Transactions on Automatic Control* 33: 955-961

Niedzwiecki M. (1994). Identification of time-varying systems with abrupt parameter changes. *Automatica* 30: 447-459

Niedzwiecki M. (2000). *Identification of Time-Varying Process*. New York: Wiley.

Ozaki T, Sosa PV, Haggan-Ozaki V. (1999). Reconstructing the nonlinear dynamics of epilepsy data using nonlinear time series analysis. *J. Signal Process* 3

Pachori RB, Sircar P. (2008). EEG signal analysis using FB expansion and second-order linear TVAR process. *Signal Processing* 88: 415-420

Palus M. (1996). Nonlinearity in normal human EEG: Cycles, temporal asymmetry, nonstationarity and randomness, not chaos. *Biological Cybernetics* 75: 389-396

Parikh D, Ahmed N, Stearns S. (1980). An adaptive lattice algorithm for recursive filters. *IEEE Transactions on Acoustics Speech and Signal Processing* 28: 110-111

Park DJ, Jun BE. (1992). Selfperturbing Recursive Least-Squares Algorithm with Fast Tracking Capability. *Electronics Letters* 28: 558-559

Peng H, Nakano K, Shioya H. (2007). Nonlinear predictive control using neural nets-based local linearization ARX model-stability and industrial application. *IEEE Transactions on Control Systems Technology* 15: 130-143

Proakis JG, Manolakis DG. (1996). *Digital signal processing principles, algorithms and applications, 3rd ed.,*. Prentice Hall, New Jersey.

Quiroga RQ. (1998). *Quantitative analysis of EEG signals: Time-Frequency methods and Chaos Theory*. Medical University of Lübeck - Germany. 145 pp.

Quiroga RQ, Kraskov A, Kreuz T, Grassberger P. (2002). Performance of different synchronization measures in real data: A case study on electroencephalographic signals. *Physical Review E* 65

Rechtschaffen A, Kales A. (1968). *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects*. University of California, Los Angeles. , US Dept. of Health, Education, and Welfare Bethesda, Md.

Rissanen J. (1978). Modelling by shortest data description. *Automatica* 14: 465-471

Sato JR, Junior EA, Takahashi DY, de Maria Felix M, Brammer MJ, Morettin PA. (2006). A method to produce evolving functional connectivity maps during the course of an fMRI experiment using wavelet-based time-varying Granger causality. *NeuroImage* 31: 187-196

Schilling RJ, Carroll JJ, Al-Ajlouni AF. (2001). Approximation of nonlinear systems with

radial basis function neural networks. *IEEE Transactions on Neural Networks* 12: 1-15

Schreiber T, Schmitz A. (1996). Improved surrogate data for nonlinearity tests. *Physical Review Letters* 77: 635-638

Schreiber T, Schmitz A. (1997). Discrimination power of measures for nonlinearity in a time series. *Physical Review E* 55: 5443-5447

Schwarz G. (1978). Estimating Dimension of a Model. *Annals of Statistics* 6: 461-464

Shynk JJ. (1992). Frequency-domain and multirate adaptive filtering. *IEEE Signal Processing Magazine* 9: 14-37

Sripada NR, Fisher DG. (1987). Improved Least-Squares Identification. *International Journal of Control* 46: 1889-1913

Subasi A. (2007). Selection of optimal AR spectral estimation method for EEG signals using Cramer-Rao bound. *Computers in Biology and Medicine* 37: 183-194

Subasi A, Ercelebi E. (2005). Classification of EEG signals using neural network and logistic regression. *Computer Methods and Programs in Biomedicine* 78: 87-99

Sugihara G, May RM. (1990). Nonlinear forecasting as a way of distinguishing chaos from measurement error in time-series. *Nature* 344: 734-741

Toplis B, Pasupathy S. (1988). Tracking Improvements in Fast RLS Algorithms Using A Variable Forgetting Factor. *IEEE Transactions on Acoustics Speech and Signal Processing* 36: 206-227

Tsatsanis MK, Giannakis GB. (1993). Time-varying system-identification and model validation using wavelets. *IEEE Transactions on Signal Processing* 41: 3512-3523

Verdes P. (2005). Assessing causality from multivariate time series. *Physical Review E* 72

Wei HL, Billings SA. (2002). Identification of time-varying systems using multiresolution wavelet models. *International Journal of Systems Science* 33: 1217-1228

Wei HL, Billings SA. (2004). A unified wavelet-based modelling framework for non-linear system identification: the WANARX model structure. *International Journal of Control* 77: 351-366

Wei HL, Billings SA. (2006a). An efficient nonlinear cardinal B-spline model for high tide forecasts at the Venice Lagoon. *Nonlinear Processes in Geophysics* 13: 577-584

Wei HL, Billings SA. (2006b). Long term prediction of non-linear time series using

multiresolution wavelet models. *International Journal of Control* 79: 569-580

Wei HL, Billings SA. (2007). Feature subset selection and ranking for data dimensionality reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29: 162-166

Wei HL, Billings SA. (2008). Model structure selection using an integrated forward orthogonal search algorithm assisted by squared correlation and mutual information. *International Journal of Modelling, Identification and Control* 3: 341-356

Wei HL, Billings SA. (2009). Improved model identification for non-linear systems using a random subsampling and multifold modelling (RSMM) approach. *International Journal of Control* 82: 27-42

Wei HL, Billings SA, Balikhin M. (2004a). Analysis of the geomagnetic activity of the D-st index and self-affine fractals using wavelet transforms. *Nonlinear Processes in Geophysics* 11: 303-312

Wei HL, Billings SA, Balikhin MA. (2006). Wavelet based non-parametric NARX models for nonlinear input-output system identification. *International Journal of Systems Science* 37: 1089-1096

Wei HL, Billings SA, Liu J. (2004b). Term and variable selection for non-linear system identification. *International Journal of Control* 77: 86-110

Wei HL, Billings SA, Liu JJ. (2010). Time-varying parametric modelling and time-dependent spectral characterisation with applications to EEG signals using multiwavelets. *International Journal of Modelling, Identification and Control* 9: 215-224

Wei HL, Lang ZQ, Billings SA. (2008). Constructing an overall dynamical model for a system with changing design parameter properties. *International Journal of Modelling, Identification and Control* 5: 93-104

Wei HL, Zheng Y, Pan Y, Coca D, Li LM, et al. (2009). Model Estimation of Cerebral Hemodynamics Between Blood Flow and Volume Changes: A Data-Based Modelling Approach. *IEEE Transactions on Biomedical Engineering* 56: 1606-1616

Wellstead P, Zarrop MB. (1991). *Self-tuning systems: control and signal processing*. John Wiley & Sons, Inc. New York, NY, USA.

West M, Prado R, Krystal AD. (1999). Evaluation and comparison of EEG traces: Latent structure in nonstationary time series. *Journal of the American statistical association*: 1083-1095

Wigner E. (1932). On the quantum correction for thermodynamic equilibrium. *Physical Review* 40: 0749-0759

Yang JN, Lei Y, Pan S, Huang N. (2003). System identification of linear structures based on Hilbert-Huang spectral analysis. Part 1: normal modes. *Earthquake Engineering & Structural Dynamics* 32: 1443-1467

Zheng YJ, Lin ZP. (2003). Recursive adaptive algorithms for fast and rapidly time-varying state. *IEEE Transactions on Circuits and Systems Ii-Analog and Digital Signal Processing* 50: 602-614

Zhou SM, Gan JQ, Sepulveda F. (2008). Classifying mental tasks based on features of higher-order statistics from EEG signals in brain-computer interface. *Information Sciences* 178: 1629-1640

Zhu QM, Billings SA. (1996). Fast orthogonal identification of nonlinear stochastic models and radial basis function neural networks. *International Journal of Control* 64: 871-886

Zou R, Chon KH. (2004). Robust algorithm for estimation of time-varying transfer functions. *IEEE Transactions on Biomedical Engineering* 51: 219-228

Zou R, Wang HL, Chon KH. (2003). A robust time varying identification algorithm using basis functions. *Annals of Biomedical Engineering* 31: 840-853